A simple dynamic model explains the diversity of island birds worldwide

Valente, Luis; Phillimore, Albert B; Melo, Martim; Warren, Ben H; Clegg, Sonya M; Havenstein, Katja; Tiedemann, Ralph; Illera, Juan Carlos; Thébaud, Christophe; Aschenbach, Tina

Published in:
Nature

DOI:
10.1038/s41586-020-2022-5

Publication date:
2020

Article

A simple dynamic model explains the diversity of island birds worldwide

https://doi.org/10.1038/s41586-020-022-5

Received: 19 March 2019

Accepted: 22 January 2020

Published online: 19 February 2020

Check for updates

Luis Valente[1,2,3,4]✉, Albert B. Phillimore[5], Martim Melo[6,7,8], Ben H. Warren[9], Sonya M. Clegg[10,11], Katja Havenstein[4], Ralph Tiedemann[4], Juan Carlos Illera[12], Christophe Thébaud[13], Tina Aschenbach[1] & Rampal S. Etienne[3]

Colonization, speciation and extinction are dynamic processes that influence global patterns of species richness[1–6]. Island biogeography theory predicts that the contribution of these processes to the accumulation of species diversity depends on the area and isolation of the island[7,8]. Notably, there has been no robust global test of this prediction for islands where speciation cannot be ignored[9], because neither the appropriate data nor the analytical tools have been available. Here we address both deficiencies to reveal, for island birds, the empirical shape of the general relationships that determine how colonization, extinction and speciation rates co-vary with the area and isolation of islands. We compiled a global molecular phylogenetic dataset of birds on islands, based on the terrestrial avifaunas of 41 oceanic archipelagos worldwide (including 596 avian taxa), and applied a new analysis method to estimate the sensitivity of island-specific rates of colonization, speciation and extinction to island features (area and isolation). Our model predicts—with high explanatory power—several global relationships. We found a decline in colonization with isolation, a decline in extinction with area and an increase in speciation with area and isolation. Combining the theoretical foundations of island biogeography[7,8] with the temporal information contained in molecular phylogenies[10] proves a powerful approach to reveal the fundamental relationships that govern variation in biodiversity across the planet.

A key feature of global diversity is the tendency for some areas to harbour many more species than others[7,8]. Uncovering the drivers and regulators of spatial differences in diversity of simple systems such as islands is a crucial step to understanding the global distribution of species richness. The two most prominent biodiversity patterns in fragmented or isolated environments worldwide are the increase in species richness with area and the decline in species richness with isolation[8,11–14]. In their theory of island biogeography, MacArthur and Wilson proposed how the processes of colonization and extinction could explain these patterns[7,8]. They argued that the rates of these processes are determined by the geographical context: colonization decreases with isolation and extinction decreases with area[7,8]. They also suggested that rates of formation of island endemic species through in situ speciation increase with island isolation and area[8]. Despite an abundance of studies over five decades that support the general patterns predicted by MacArthur and Wilson[2,15–18], tests of predictions regarding the dependence of the underlying processes—colonization, speciation and extinction—on island geographical context (area and isolation) are few in number, and are either restricted in temporal, geographical or taxonomic scope[5,19,20], or seek to infer speciation rates in the absence of data on the phylogenetic relationships among species[2,16]. As a result, there has been no robust and powerful test of MacArthur and Wilson's predictions on a global scale, and the effect of area and isolation on biogeographical processes acting on macroevolutionary timescales remains largely unexplored.

Here we expand on approaches that leverage the information in time-calibrated molecular phylogenies of insular species[1,10,21,22] to determine how the processes of colonization, speciation and extinction are influenced by area and isolation. The dynamic stochastic model DAISIE[10] (dynamic assembly of islands through speciation, immigration and extinction) can accurately estimate maximum-likelihood rates of colonization, extinction and speciation rates (CES rates) from branching times (colonization times and any in situ diversification events) and endemicity status of species that results from one or multiple independent colonizations of a given island system (for example, all native terrestrial birds on an archipelago)[10]. This method can also detect the presence or absence of diversity dependence in rates of colonization and speciation, by estimating a carrying capacity (upper bound to the

[1]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany. [2]Naturalis Biodiversity Center, Leiden, The Netherlands. [3]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands. [4]Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. [5]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. [6]Museu de História Natural e da Ciência da Universidade do Porto, Porto, Portugal. [7]Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), InBio, Laboratório Associado, Universidade do Porto, Vairão, Portugal. [8]FitzPatrick Institute, DST-NRF Centre of Excellence, University of Cape Town, Cape Town, South Africa. [9]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, UA, Paris, France. [10]Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK. [11]Environmental Futures Research Institute, Griffith University, Brisbane, Queensland, Australia. [12]Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, Mieres, Spain. [13]Unité Mixte de Recherche 5174, CNRS-IRD-Paul Sabatier University, Toulouse, France. ✉e-mail: luis.valente@naturalis.nl

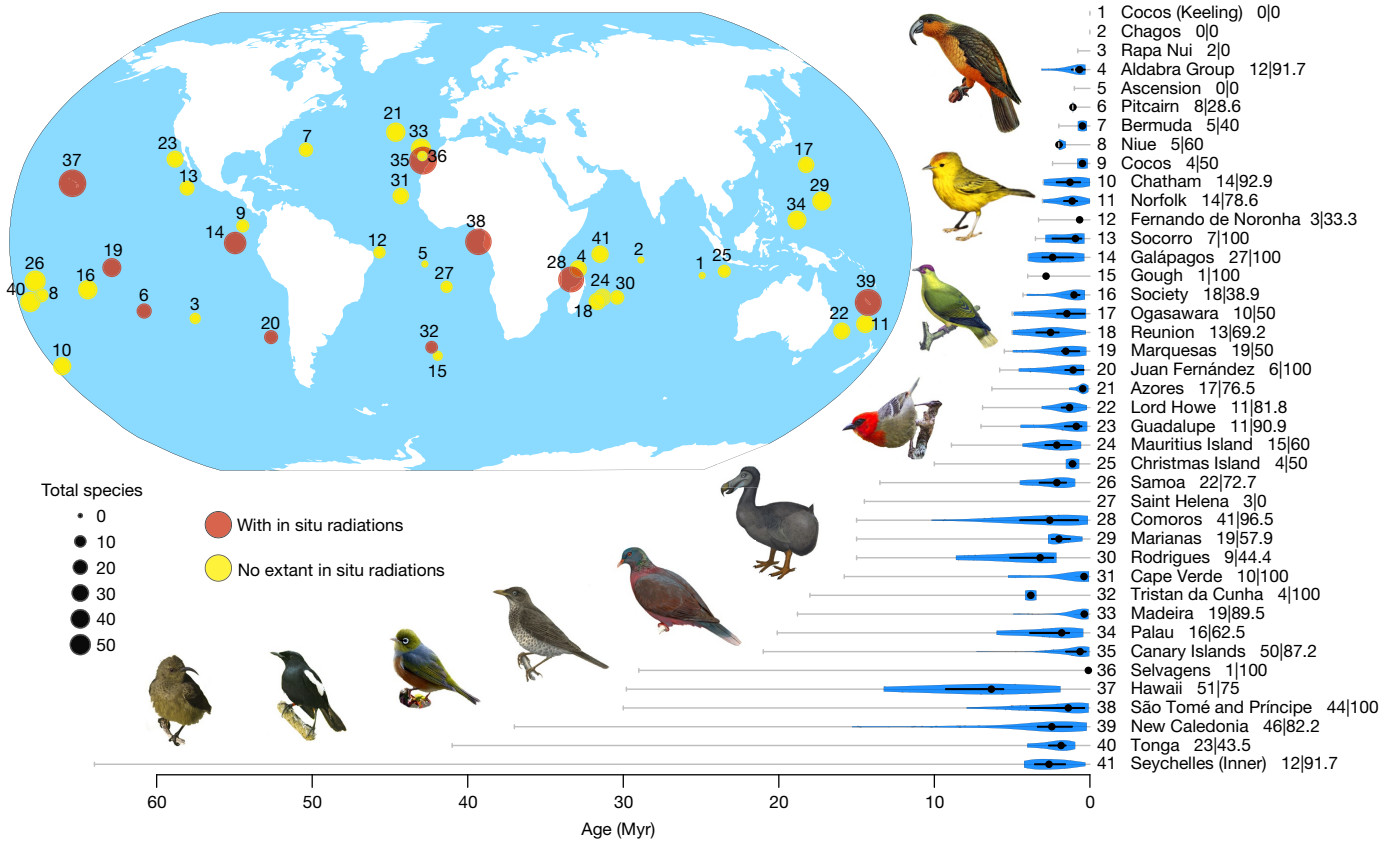


Fig. 1 | Archipelago and island bird colonization time data. Circles show the number of species that belong to our focal group (both extinct and extant) found in each archipelago (at the time of human arrival). Numbers on the map correspond to numbers to the left of the archipelago name. Numbers to the right of the archipelago name indicate the number of species from our focal assemblage on the archipelago | the percentage of species sampled in the phylogenetic trees. Even species not sampled in the trees are accounted for by including them as missing species that could have colonized at any time since the emergence of the archipelago. Colonization times plot: grey horizontal lines indicate archipelago ages (Extended Data Table 1). Violin plots (blue) show the kernel density of the distribution of times of colonization of bird species in each archipelago, obtained from the phylogenetic trees. Thick black lines inside violin plots indicate the interquartile distance; thin black lines indicate the 95% confidence interval; black dots indicate the median. Archipelagos with no violin plot or dots are cases for which no species of our focal assemblage were present at the time of human arrival, or none were sampled using molecular data. Birds from left to right: Seychelles sunbird, Seychelles magpie robin, silvereye, Príncipe thrush, laurel pigeon, dodo (extinct), Mauritius fody, red-moustached fruit dove (extinct), Galápagos warbler and Norfolk kaka (extinct). Bird images used with permission from: C. Baeta (Príncipe thrush), P. Cascão (Galápagos warbler), M. Hammers (Seychelles sunbird and magpie robin), J. Hume (dodo), D. Shapiro (Mauritius fody) and J. Varela (laurel pigeon). There are no in situ radiations in the Mascarenes (Mauritius, Reunion and Rodrigues) because we treat the islands as separate entities (but see 'Sensitivity to archipelago selection and isolation metrics' in the Methods). Myr, million years.

number of species in an island system). Here we extend DAISIE to estimate the hyperparameters that control the shape of the relationships between CES rates, and the area and isolation of islands worldwide.

The accurate estimation of fundamental island biogeographical relationships requires suitable data from many archipelagos, but divergence-dated phylogenies of complete communities on islands remain scarce. Hence, we produced new dated molecular phylogenies for the terrestrial avifaunas of 41 archipelagos worldwide. Here we refer to both true archipelagos (composed of multiple islands) and isolated insular units that consist of single islands (for example, Saint Helena) as 'archipelago'. For each archipelago, we compiled avian taxon lists (excluding introduced, marine, migratory and aquatic species, as well as birds of prey, rails and nocturnal birds; see Methods) and collected physical data (Fig. 1 and Supplementary Data 1, 2). We use archipelagos as our insular unit, because the high dispersal abilities of birds within archipelagos suggest that, for birds, archipelagos can be considered equivalent to single islands for less dispersive taxa[23], and because archipelagos constitute the most-appropriate spatiotemporal unit for framing analyses of biodiversity patterns at a large scale[2,24,25]. We extracted colonization and speciation times for each archipelago from the phylogenetic trees, producing a 'global dataset' for the 41 archipelagos, which includes the complete extant avifauna of each archipelago, plus all species known to have become extinct due to anthropogenic causes. The dataset comprises 596 insular taxa from 491 species. The phylogenies revealed a total of 502 archipelago colonization events and 26 independent in situ 'radiations' (cases in which diversification has occurred within an archipelago), which ranged in size from 2 to 33 species (the Hawaiian honeycreepers being the largest clade). The distribution of colonization times is summarized in Fig. 1 and the full dataset is provided in Supplementary Data 1.

Our extension of the DAISIE framework enables us to estimate hyperparameters that control the relationship between archipelago area and isolation, and archipelago-specific local CES rates, that is, rates of colonization, cladogenesis (within-archipelago speciation that involves in situ lineage splitting), anagenesis (within-archipelago speciation by divergence from the mainland without in situ lineage splitting), natural extinction rates and carrying capacity. We tested the hypothesis that area and distance from the nearest mainland have an effect on the specific CES rates, and, in cases in which a significant effect was identified, estimated its shape and scaling. We developed a set of a priori

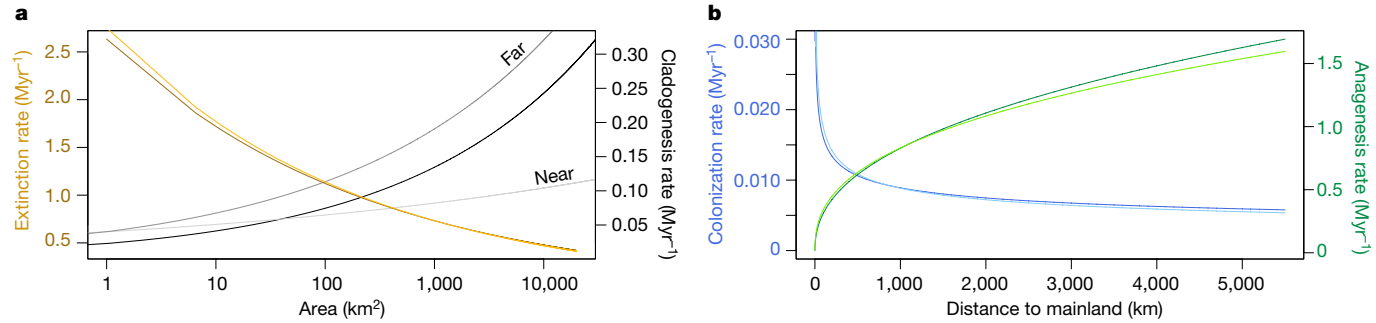


Fig. 2 | Estimated relationships between island area and isolation, and local island biogeography parameters. Isolation was measured as the distance to the nearest mainland. Relationships shown are based on the maximum likelihood global hyperparameters of the best models (equations describing the relationships are provided in Supplementary Table 1). Darker lines, M14 model; lighter lines, M19 model. Under the M14 model, the cladogenesis rate depends only on the area. Under the M19 model, the cladogenesis rate increases with both area and distance to the nearest mainland, and thus lines for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown. See Extended Data Fig. 1 for the relationship of cladogenesis with both area and distance under the M19 model.

models (Supplementary Table 1) in which the CES rates are power-law functions of archipelago features. Area has been proposed to have a positive effect on cladogenesis and carrying capacity[3,5,8] and a negative effect on extinction rates[8,26]. Archipelago isolation is hypothesized to reduce colonization rates[7] and increase anagenesis rates[27]. Models that include or exclude diversity dependence in rates of colonization and cladogenesis[10] (that is, estimating a carrying capacity parameter) were compared. We also considered a set of post hoc models with alternative shapes for the relationships (post hoc power and post hoc sigmoid models; Methods and Supplementary Table 1).

We fitted a set of 28 candidate models to the global dataset using maximum likelihood (Supplementary Table 2). The shape of the relationship of CES rates with area and distance for the two best models is shown in Fig. 2. Under the preferred a priori model (lowest value of the Bayesian information criterion; M14, eight parameters) colonization rates decline with archipelago isolation (exponent of the power law = −0.25 (95% confidence interval = −0.17–−0.34)) and extinction rate decreases with area (scaling = −0.15 (−0.11–−0.18)). Rates of cladogenesis increase with area (scaling = 0.26 (0.13–0.37)), while anagenesis increases with isolation (scaling = 0.42 (0.24–0.61)). The preferred post hoc model (M19, eight parameters) was also the preferred model overall and differs qualitatively from the preferred a priori model M14 only in the cladogenesis function. In the M14 model, cladogenesis is solely a function of area, whereas in the M19 model cladogenesis depends interactively and positively on both area and distance from the nearest mainland, such that the cladogenesis–area relationship is steeper for more isolated archipelagos (Fig. 2 and Extended Data Fig. 1). In addition, we found no evidence for diversity dependence, as the carrying capacity (K) was estimated to be much larger than the number of species on the island and models without a K parameter (no upper bound to diversity), such as M14 and M19, performed better than models that included this parameter (Supplementary Table 2). We also tested whether the inclusion of a combination of true archipelagos and single islands in our dataset could have affected our results, for example if opportunities for allopatric speciation are higher when an area is subdivided into multiple islands[28]. We repeated analyses in which single island units were excluded and found that the same model (M19) is preferred with similar parameter estimates. We therefore discuss only the results for the main dataset (including both single islands and true archipelagos). Our results are robust to uncertainty in colonization and branching times (see 'Sensitivity to alternative divergence times and tree topologies' in the Methods).

A parametric bootstrap analysis of the two preferred models (M14 and M19) demonstrated that the method is able to recover hyperparameters with high precision and little bias (Extended Data Fig. 2). To test the significance of the relationships between area, isolation and CES rates, we conducted a randomization test on the global dataset with reshuffled areas and distances. This test estimated the exponent hyperparameters as zero in most reshuffled cases (that is, no effect of area or isolation was detected; Extended Data Fig. 3), confirming that it is the observed relationships between diversity and archipelago characteristics that generate our parameter estimates.

To assess model fit, we simulated archipelago communities under the best model (M19) and found that—for most archipelagos—the observed diversity metrics (the numbers of species, cladogenetic species and colonizations) were similar to the expected numbers, with some exceptions; for example, diversity was underestimated for Comoros and São Tomé and Príncipe (Fig. 3 and Extended Data Fig. 4). The ability of the model to explain observed values (total species, pseudo-R^2 = 0.72; cladogenetic species, pseudo-R^2 = 0.52; colonizers, pseudo-R^2 = 0.60) was very high considering the model includes only 8 parameters (at least 12 parameters would be needed if each rate depended on area and isolation, and at least 164 parameters if each archipelago was allowed to have its own parameters) and was able to explain multiple diversity metrics. This represents a very large proportion of the explanatory power that would be expected to be obtained for data generated under the preferred model (Extended Data Fig. 5). Simulations under the best model reproduced the classic observed relationships between area, distance and diversity metrics (Fig. 4).

Our approach reveals the empirical shape of fundamental biogeographical relationships that have previously been difficult to estimate. In agreement with recent studies[2,29], we found strong evidence for a decline in the rates of colonization with isolation and in the rates of extinction with area, confirming two of the key assumptions of island biogeography theory[7]. The colonization–isolation effect was detected despite the fact that the decline in avian species richness with distance from the nearest mainland in our empirical data was not as pronounced as in other less-mobile taxa[4,11], revealing that isolation is a clear determinant of the probability of immigration and the successful establishment of populations even in a highly dispersive group such as birds. The extinction–area relationship has been a fundamental empirical generalization in conservation ecology (for example, for the design of protected areas[30]); here we were able to characterize the shape of this dependence at the global spatial scale and macro-evolutionary timescale.

We provide insights into the scaling of speciation with area and isolation. In contrast to previous studies on within-island speciation, which have suggested the existence of an area below which cladogenesis does not take place on single islands[5], we do not find evidence for such an area threshold at the archipelago level and, under our model, speciation is predicted to be non-zero even in small areas. In addition, our post

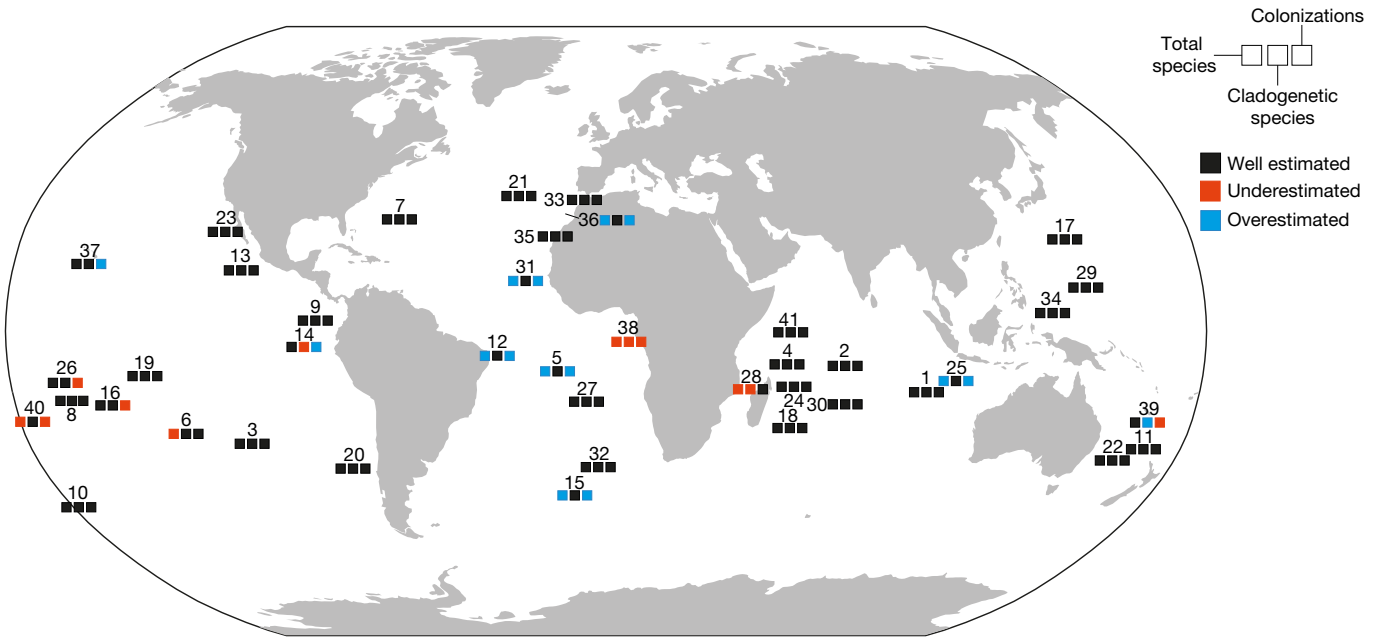


Fig. 3 | Goodness of fit of the preferred model (M19). The map identifies whether the diversity metrics were well estimated (the empirical value matches the 95% confidence interval of simulations), underestimated (the empirical value is higher than the 95% confidence interval) or overestimated (the empirical value is lower than the 95% confidence interval). Intervals are based on 1,000 simulations of each archipelago (Extended Data Fig. 4). Numbers on the map indicate the archipelagos described in Fig. 1.

hoc finding that rates of cladogenesis increase through an interactive effect of both island size and distance from the nearest mainland (Fig. 2 and Extended Data Fig. 1) provides a mechanism that limits radiations to archipelagos that are both large and remote[6,27]. Why this interaction exists requires further investigation, but one possibility is that unsaturated niche space provides greater opportunities for diversification[6]. In addition to the effects of physical features on cladogenesis, we found that rates of anagenesis increase with island isolation. While impressive insular radiations tend to receive the most attention from evolutionary biologists (for example, Darwin's finches or Hawaiian

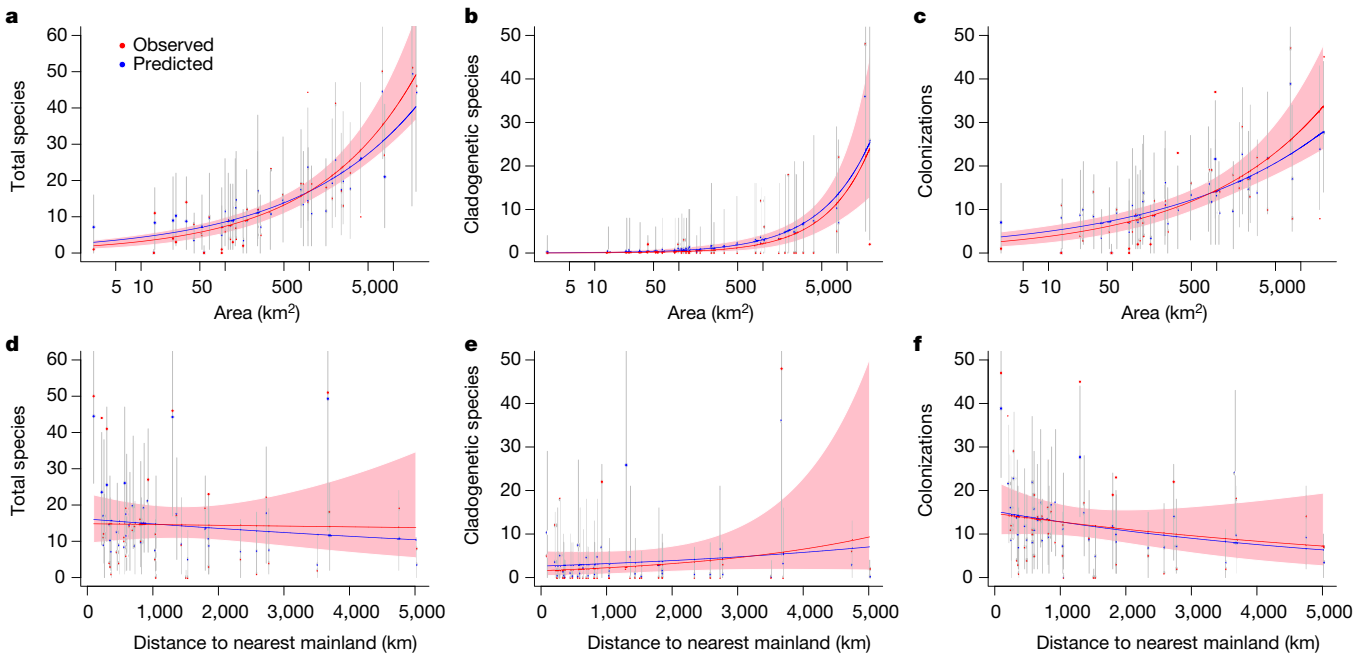


Fig. 4 | Observed and predicted island diversity–area and island diversity–distance relationships. Grey vertical lines show the 95% confidence intervals across 1,000 datasets simulated for each of the 41 archipelagos assuming the M19 model. Blue points indicate the mean values of the simulations; the blue line indicates the fitted line for the simulated data; red points are the observed values in the empirical data; the red line shows the fitted line for the empirical data; the red shaded area is the 95% confidence interval of the predicted relationship for the empirical data. **a**–**c**, Relationships between island diversity and area. **a**, Total number of species. **b**, Cladogenetic species. **c**, Number of colonizations. **d**–**f**, Relationships between island diversity and distance of the island to the mainland. **d**, Total number of species. **e**, Cladogenetic species. **f**, Number of colonizations.

honeycreepers), our phylogenies revealed that the majority of endemic birds in our dataset in fact display an anagenetic pattern (at the time of human arrival, 231 out of 350 endemic species had no extant sister taxa on the archipelago and there were only 26 extant in situ radiations). The positive effect of archipelago isolation on rates of anagenesis that we estimate suggests that this fundamental but overlooked process is impeded by high levels of movement between island and mainland populations.

A variety of global patterns of biodiversity have been described—from small islands and lakes, to biomes and continents—but the processes that underpin these patterns remain to be explored. Our simulations using parameters estimated from data were able to reproduce the classic global patterns of island biogeography across 41 archipelagos (Fig. 4). This advances our understanding of macroscale biology, by providing missing links between local processes, environment and global patterns. More than half a century after the seminal work of MacArthur and Wilson[7], we now have the data and tools to go beyond statistical descriptions of diversity patterns, enabling us to quantify community-level processes that have long been unclear.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2022-5.

1. Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science* **294**, 1522–1524 (2001).
2. Triantis, K. A., Economo, E. P., Guilhaumon, F. & Ricklefs, R. E. Diversity regulation at macroscales: species richness on oceanic archipelagos. *Glob. Ecol. Biogeogr.* **24**, 594–605 (2015).
3. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
4. Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. Global diversity of island floras from a macroecological perspective. *Ecol. Lett.* **11**, 116–127 (2008).
5. Losos, J. B. & Schluter, D. Analysis of an evolutionary species–area relationship. *Nature* **408**, 847–850 (2000).
6. Gillespie, R. G. & Baldwin, B. G. in *The Theory of Island Biogeography Revisited* (eds Losos, J. & Ricklefs, R. E.) 358–387 (Princeton Univ. Press, 2010).
7. MacArthur, R. H. & Wilson, E. O. An equilibrium theory of insular zoogeography. *Evolution* **17**, 373–387 (1963).
8. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
9. Warren, B. H. et al. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur–Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
10. Valente, L. M., Phillimore, A. B. & Etienne, R. S. Equilibrium and non-equilibrium dynamics simultaneously operate in the Galápagos islands. *Ecol. Lett.* **18**, 844–852 (2015).
11. Lomolino, M. V. Species–area and species–distance relationships of terrestrial mammals in the Thousand Island Region. *Oecologia* **54**, 72–75 (1982).
12. Diamond, J. M. Biogeographic kinetics: estimation of relaxation times for avifaunas of southwest Pacific islands. *Proc. Natl Acad. Sci. USA* **69**, 3199–3203 (1972).
13. Whittaker, R. J. & Fernandez-Palacios, J. M. *Island Biogeography: Ecology, Evolution, and Conservation* (Oxford Univ. Press, 2007).
14. Matthews, T. J., Rigal, F., Triantis, K. A. & Whittaker, R. J. A global model of island species–area relationships. *Proc. Natl Acad. Sci. USA* **116**, 12337–12342 (2019).
15. Weigelt, P., Steinbauer, M. J., Cabral, J. S. & Kreft, H. Late Quaternary climate change shapes island biodiversity. *Nature* **532**, 99–102 (2016).
16. Lim, J. Y. & Marshall, C. R. The true tempo of evolutionary radiation and decline revealed on the Hawaiian archipelago. *Nature* **543**, 710–713 (2017).
17. Cabral, J. S., Weigelt, P., Kissling, W. D. & Kreft, H. Biogeographic, climatic and spatial drivers differentially affect α-, β- and γ-diversities on oceanic archipelagos. *Proc. R. Soc. B* **281**, 20133246 (2014).
18. Matthews, T. J., Guilhaumon, F., Triantis, K. A., Borregaard, M. K. & Whittaker, R. J. On the form of species–area relationships in habitat islands and true islands. *Glob. Ecol. Biogeogr.* **25**, 847–858 (2016).
19. Simberloff, D. S. & Wilson, E. O. Experimental zoogeography of islands: the colonization of empty islands. *Ecology* **50**, 278–296 (1969).
20. Russell, G. J., Diamond, J. M., Reed, T. M. & Pimm, S. L. Breeding birds on small islands: island biogeography or optimal foraging? *J. Anim. Ecol.* **75**, 324–339 (2006).
21. Rabosky, D. L. & Glor, R. E. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Natl Acad. Sci. USA* **107**, 22178–22183 (2010).
22. Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* **23**, 619–630 (2008).
23. Kisel, Y. & Barraclough, T. G. Speciation has a spatial scale that depends on levels of gene flow. *Am. Nat.* **175**, 316–334 (2010).
24. Triantis, K., Whittaker, R. J., Fernández-Palacios, J. M. & Geist, D. J. Oceanic archipelagos: a perspective on the geodynamics and biogeography of the World's smallest biotic provinces. *Front. Biogeogr.* **8**, e29605 (2016).
25. Santos, A. M. C. et al. Are species–area relationships from entire archipelagos congruent with those of their constituent islands? Glob. *Ecol. Biogeogr.* **19**, 527–540 (2010).
26. Ricklefs, R. E. & Lovette, I. J. The roles of island area per se and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. *J. Anim. Ecol.* **68**, 1142–1160 (1999).
27. Rosindell, J. & Phillimore, A. B. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* **14**, 552–560 (2011).
28. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
29. Keil, P. et al. Spatial scaling of extinction rates: theory and data reveal nonlinearity and a major upscaling and downscaling challenge. *Glob. Ecol. Biogeogr.* **27**, 2–13 (2016).
30. Wilcox, B. A. & Murphy, D. D. Conservation strategy: the effects of fragmentation on extinction. *Am. Nat.* **125**, 879–887 (1985).

Methods

Archipelago selection

We focus on oceanic islands, that is, volcanic islands that have never been connected to any other landmass in the past. We also include the Granitic Inner Seychelles, even though these islands have a continental origin, because they have been separated from other landmasses for a very long period of time (64 million years)[31] and can be considered quasi-oceanic, as all extant avian species originated in much more recent times. The 41 archipelagos chosen are located in the Atlantic, Indian and Pacific Oceans, with latitudes between 45° north and south. Islands within these archipelagos are separated by a maximum of 150 km. The sole exceptions are the Azores and Hawaii, two very isolated systems where the distances between some islands exceed this value. The shape files used to plot the maps of Figs. 1, 3 were obtained from a previous study[32].

Physical and geological data

Full archipelago data are provided in Supplementary Data 2 and Extended Data Table 1. We obtained data on the total contemporary landmass area for each archipelago. For our isolation metric, we computed the minimum round Earth distance to the nearest mainland (D_m) in km using Google Earth. We considered 'nearest mainland' to be the nearest probable source of colonists (but see 'Sensitivity to archipelago selection and isolation metrics' for different isolation metrics). This is the nearest continent except for island groups that were closer to Madagascar, New Guinea or New Zealand than to the continent, in which case we assigned these large continent-like islands as the mainland. This is supported by our phylogenetic data—for example, many Indian Ocean island taxa have closest relatives on Madagascar rather than mainland Africa.

Island palaeo-areas and past archipelago configurations have been shown to be better predictors of endemic insular diversity than contemporary area[15,33]. By contrast, island total native and non-endemic richness is better predicted by present island characteristics[15,33]. As insufficient data on island ontogeny was available (that is, describing the empirical area trajectories from island birth to present), we analysed contemporary area and isolation as these are currently the most appropriate units for our dataset.

We conducted an extensive survey of the literature and consulted geologists to obtain the geological ages for each archipelago (Extended Data Table 1), treating the age of the oldest currently emerged island as an upper bound for colonization. Islands may have been submerged and have emerged multiple times and we consider the age of the last known emergence. For the Aldabra Group we used an age older than the published estimate. The current estimated age of re-emergence of Aldabra is 0.125 million years[34], but 9 out of 12 Aldabra colonization events in our dataset are older, suggesting that the archipelago was not fully submerged before this and may have been available for colonization for a longer period. Therefore, for Aldabra we used an older upper bound of 1 million years for colonization, although we acknowledge that the mitochondrial markers used for dating may not provide sufficient resolution at the shallow temporal scale of the published age. For Hawaii, the colonization times that we obtained for more than half of the colonization events were older than the age of the current high islands that is often used as a maximum age for colonization (around 5 million years). Therefore, instead of this age, we used the much older estimate of 29.8 million years of the Kure Atoll[35] to account for currently submerged or very low-lying Hawaiian Islands that could have received colonists in the past. For Bermuda and Marianas, we could not find age estimates in the literature, and we therefore consulted geologists to obtain these (P. Hearty, R. Stern and M. Reagan, personal communication; Extended Data Table 1).

Island avifaunas

Our sampling focused on native resident terrestrial birds and we considered only birds that colonize by chance events (for example, hurricanes or rafts). We thus excluded marine and migratory species, because they are capable of actively colonizing an island at a much higher rate. We focused on songbird-like and pigeon-like birds, which constitute the majority of terrestrial (land-dwelling) birds on islands. Following a precedent set by previous work[10,27,36], we included only species from the same trophic level (in the spirit of MacArthur and Wilson's model): we excluded aquatic birds, birds of prey, rails (many are flightless or semi-aquatic) and nightjars (nocturnal). We also excluded introduced and vagrant species. Including species such as rails and owls (which are components of many island avifaunas) would have led to a higher estimate of the product of colonization rate and mainland pool size due to a larger mainland pool, and potentially to higher estimated rates of anagenesis (many owl or rail species are island endemics with no close relatives on the islands).

For the focal avian groups, we compiled complete taxon lists for each of the 41 archipelagos based on recent checklists from Avibase (http://avibase.bsc-eoc.org), which we cross-checked with the online version of the *Handbook of the Birds of the World* (HBW)[37]. We followed the HBW's nomenclature and species assignations, except for 12 cases in which our phylogenetic data disagree with HBW's scheme (noted in the column 'Taxonomy' of Supplementary Data 1). For example, in 11 cases phylogenetic trees support raising endemic island subspecies to species status (we sampled multiple samples per island taxon and outgroup, and the island individuals form a reciprocally monophyletic well-supported clade), and for these taxa we decided it was more appropriate to use a phylogenetic species concept so as not to underestimate endemicity and rates of speciation (Supplementary Data 1). We re-ran DAISIE analyses using HBW's classification and found that the maximum-likelihood parameters are very similar and thus we report only the results using the scheme based on the phylogenies produced for this study.

For each bird species found on each archipelago, we aimed to sample sequence data for individuals on the archipelago and the closest relatives outside the archipelago (outgroup taxa). Our sampling success per archipelago is shown in Fig. 1 and Extended Data Table 1.

Extinct species

We do not count extinctions with anthropogenic causes as influencing the natural background rate of extinction. Therefore, we explicitly include species for which there is strong evidence that they have been extirpated by humans. We treat taxa extirpated on an archipelago by humans as though they had survived in that archipelago until the present following our previously published approach[38].

We identified anthropogenic extinctions based on published data[39–46] and personal comments (J. A. Alcover and J. C. Rando on unpublished Macaronesian taxa; F. Sayol and S. Faurby). We include the species present on the islands that belong to our archipelago definition as described in Supplementary Data 2. We excluded largely hypothetical accounts or pre-Holocene fossils that greatly predate human arrival. Our dataset accounts for 153 taxa that were present on first human contact and have gone extinct since, probably because of human activities including the introduction of invasive species by humans. To our knowledge, 71 of these taxa have previously been sequenced using ancient DNA or belong to clades present in our trees, and we were thus able to include them in the phylogenetic analyses as regular data ($n = 54$), or as missing species by adding them as unsampled species to a designated clade ($n = 17$). For the remaining 82 extinct taxa, sequences were not available and we were unable to obtain samples and to allocate them to clades. We assume that these taxa represent extinct independent colonizations and we included them in the analyses using the 'Endemic_MaxAge' and 'Non_endemic_MaxAge' options in DAISIE, which assume that they have colonized at any given time since the birth of the archipelago (but before any in situ cladogenesis event). As an example, our dataset includes the 27 species of Hawaiian birds belonging to our focal group that are known to have gone extinct since human

colonization. Eight of these species were included using DNA data, 17 were added as missing species to their clades (14 honeycreepers and 3 *Myadestes*) and two were added using the Endemic_MaxAge option in DAISIE (*Corvus impluviatus* and *Corvus viriosus*).

Sequence data from GenBank

We conducted an extensive search of GenBank for available DNA sequences from the 596 island bird taxa that fitted our sampling criteria and from multiple outgroup taxa, using Geneious v.11[47]. The molecular markers chosen varied from species to species, depending on which marker was typically sequenced for the taxon in question, the most commonly sequenced marker was cytochrome *b*. In total, we downloaded 3,155 sequences from GenBank. For some taxa, sequences from both archipelago and close relatives from outside the archipelago were already available from detailed phylogenetic or phylogeographical analyses. In some cases, a target species had been sampled, but only from populations outside the archipelago. In other cases, the species on the archipelago had been sampled, but the sampling of the relatives outside of the archipelago was lacking or only available for distant regions, which meant a suitable outgroup was not available in GenBank. Finally, for some species there were no previous published sequences available in GenBank. GenBank accession numbers and geographical origin for the downloaded sequences are provided in the DNA matrices (https://doi.org/10.17632/vf95364vx6.1) and maximum clade credibility trees (https://doi.org/10.17632/p6hm5w8s3b.2) uploaded to Mendeley Data.

Sequence data of new samples

Sequences available in GenBank covered only 54% (269 out of 502) of the total independent colonization events. We improved the sampling by obtaining new sequences for many island taxa (n = 174 taxa) and from their close relatives from continental regions (n = 78). We obtained new samples from three sources: field trips, research collections and colleagues who contributed field samples. New samples were obtained during field trips conducted by M.M. (Gulf of Guinea and African continent); B.H.W. and C.T. (Comoros and Mayotte, Mauritius Island, Rodrigues, Seychelles); S.M.C. (New Caledonia); J.C.I. (Macaronesia, Europe and Africa) and L.V. (New Caledonia), between 1999 and 2017. Samples of individuals were captured using mist-nets or spring traps baited with larvae. Blood samples were taken by brachial venipuncture, diluted in ethanol or Queen's lysis buffer in a microcentrifuge tube. Birds were released at the point of capture. Aldabra Group samples were obtained from research collections of the Seychelles Islands Foundation. Museum samples from several Galápagos and Comoros specimens were obtained on loan from, respectively, the California Academy of Sciences and the Natural History Museum London. Additional samples from various localities (Aldabra Islands, Iberian Peninsula, Madagascar and Senegal) were provided by collaborators, as indicated in Supplementary Table 3. Sample information and GenBank accession numbers for all new specimens are provided in Supplementary Table 3.

DNA was extracted from blood, feathers and museum toe-pad samples using QIAGEN DNeasy Blood and Tissue kits (Qiagen). For museum samples, we used a dedicated ancient DNA laboratory facility at the University of Potsdam to avoid contamination. The cytochrome *b* region (1,100 base pairs) was amplified using the primers shown in Extended Data Table 2. DNA from historical museum samples was degraded and cytochrome *b* could not be amplified as a single fragment. We thus designed internal primers to sequence different overlapping fragments in a stepwise manner (Extended Data Table 2).

PCRs were set up in 25-µl total volumes including 5 µl of buffer Bioline MyTaq, 1 µl (10 mM) of each primer and 0.12 µl MyTaq polymerase. PCRs were performed with the following thermocycler conditions: initial denaturation at 95 °C for 1 min followed by 35 cycles of denaturation at 95 °C for 20 s, with an annealing temperature of 48 °C for 20 s, and extension at 72 °C for 15 s and a final extension at 72 °C for 10 min. Amplified products were purified using exonuclease I and Antarctic phosphatase, and sequenced at the University of Potsdam (Unit of Evolutionary Biology/Systematic Zoology) on an ABI PRISM 3130xl sequencer (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). We used Geneious v.11 to edit chromatograms and align sequences.

Phylogenetic analyses

To estimate times of colonization and speciation for each archipelago, we produced new divergence dated phylogenies or compiled published dated trees, to yield a total of 91 independent phylogenies (maximum clade credibility trees and posterior distribution deposited in Mendeley, https://doi.org/10.17632/p6hm5w8s3b.2) for all new trees produced for this study; the 11 previously published trees are available upon request). Information on all alignments and trees, including molecular markers, data sources, calibration methods and substitution model are provided in Extended Data Tables 3, 4 and Supplementary Table 4. The majority of alignments and phylogenies focus on a single genus, although some include multiple closely related genera or higher order clades (family, order) depending on the diversity and level of sampling of the relevant group (taxonomic scope is described in Extended Data Tables 3, 4). Most alignments include taxa from a variety of archipelagos. Alignments were based on a variety of markers, according to which marker had most often been sequenced for a given group.

For the new dating analyses conducted for this study, we created 80 separate alignments for different groups using a combination of sequences from GenBank (n = 3,155) and new sequences (n = 252) produced for this study. In some cases, we obtained DNA alignments directly from authors of previous studies and these are credited in Extended Data Table 3. Phylogenetic divergence dating analyses were performed in BEAST 2[48]. For each alignment, we performed substitution model selection in jModeltest[49] using the Bayesian information criterion (BIC). We used rates of molecular evolution for avian mitochondrial sequences, which have been shown to evolve in a clock-like manner at an average rate of around 2% per million years[50]. Molecular rate calibrations can be problematic for ancient clades, due to high levels of heterotachy in birds[51]. In addition, mitochondrial DNA saturates after about 10–20 million years, and genetic distances of more than 20% may provide limited information regarding dating[52]. Therefore, we only used molecular rate dating to extract node ages for branching events at the tips of the trees, at the species or population level (oldest colonization time in our dataset is 15.3 million years, but most are much younger). Rates of evolution were obtained from the literature and varied between different markers and taxonomic group (Supplementary Table 4). We applied the avian mitochondrial rates estimated using cytochrome *b* from a previous study[50] (but see 'Sensitivity to alternative divergence times and tree topologies' for different rates).

We applied a Bayesian uncorrelated log-normal relaxed clock model. For each analysis, we ran two independent chains of between 10 and 40 million generations, with a birth–death tree prior. We assessed convergence of chains and appropriate burn-ins with Tracer, combined runs using LogCombiner and produced maximum clade credibility trees with mean node heights in Tree Annotator. We produced a total of 80 maximum clade credibility trees.

For 11 groups (Extended Data Table 4), well-sampled and rigorously dated phylogenies were already available from recent publications, all of which conducted Bayesian divergence dating using a variety of calibration methods, including fossils and molecular rates. We obtained maximum clade credibility trees from these studies from online repositories or directly from the authors (Extended Data Table 4).

Colonization and branching times

The nodes selected in the dated trees for estimates of colonization and branching times are given for each taxon in Supplementary Data 1. Our node selection approach was as follows. For cases in which samples

representing species or populations from archipelagos formed a monophyletic clade consisting exclusively of archipelago individuals, we used the stem age of this clade as colonization time. For cases in which only one individual of the archipelago was sampled, we used the length of the tip leading to that individual, which is equivalent to the stem age. For cases in which the archipelago individuals were embedded in a clade containing mainland individuals of the same species—that is, paraphyly or polyphyly—we assumed (based on morphological characteristics) that this is due to incomplete lineage sorting of the insular and mainland lineages, and we therefore used the most recent common ancestor node of the archipelago individuals, or the crown node when the most recent common ancestor node coincides with the crown. For these later cases, using the stem would most likely have been an overestimation of the colonization time, as we assume that colonization happens from the mainland to the archipelago. For such cases, we applied the ages using the 'MaxAge' option in DAISIE, which integrates over the possible colonization times between the present and the upper bound. A robustness test of our results to node choice is given in 'Sensitivity to alternative branching times and tree topologies'.

For a total of 19 endemic taxa we could not obtain sequences, but we could allocate them to a specific island clade (for example, Hawaiian honeycreepers and solitaires). These were added as missing species to that clade. For 96 non-endemic taxa we could not obtain sequences of individuals from the archipelago, but we could obtain sequences from the same species from different regions. For these cases, we used the crown or the stem age of the species as an upper bound for the age of the colonization event, using the 'Non_endemic_MaxAge' option in DAISIE. Finally, for 124 taxa (20.8%) no sequences of individuals from the archipelago were available in GenBank and we were not able to obtain samples for sequencing from the species or from close relatives. We assumed these cases constituted independent colonizations that could have taken place any time since the origin of the archipelago and the present, and applied the 'Non_endemic_MaxAge' and 'Endemic_MaxAge' options in DAISIE with a maximum age equal to the archipelago age. DAISIE makes use of the information described above; further information has been described previously[53].

Global dataset characteristics

Data points from taxa of the same archipelago were assembled into 41 archipelago-specific datasets. These 41 datasets were in turn assembled into a single dataset (D1), which was analysed with DAISIE (D1 DAISIE R object, available in Mendeley Data https://doi.org/10.17632/sy58zbv3s2.2). This dataset (Supplementary Data 1) has a total of 596 taxa (independent colonization events plus species within radiations), covering 491 species from 203 different genera and 8 orders. All taxa were included in the analyses: not only those which we sampled in phylogenies, but also those for which sequences or phylogenies could not be obtained and which were included following the approaches described in 'Colonization and branching time'. A summary of diversity and sampling per archipelago is provided in Extended Data Table 1.

Sampling completeness

In total, we produced new sequences from 252 new individuals, comprising 90 different species from 45 different genera, covering an additional 110 colonization events that had not been sampled (that is, populations from islands for which the species had not been sampled before). For at least 12 of these 90 species, we found no previous sequences in GenBank, including island endemics from Comoros, Galápagos, Rodrigues and São Tomé (Supplementary Table 5). The new sequences from 252 individuals increase the molecular sampling for extant colonization events from 60% (223 out of 373) to 89% (332 out of 373). If we include historically extinct colonizations, we increased the molecular sampling from the existing 54% (269 out of 502) of colonization events to 75% (379 out of 502). We also substantially increased molecular sampling of continental relatives, adding 78 new individuals from the continent or islands surrounding our archipelagos, covering 43 different species. The percentage of taxa sampled in phylogenies varied widely between archipelagos (Fig. 1 and Extended Data Table 1). For 8 archipelagos (Bermuda, Fernando de Noronha, Pitcairn, Rapa Nui, Rodrigues, Saint Helena, Society Islands and Tonga) less than 50% of the species were sampled in phylogenies, and thus the majority of the species for these island groups were added with maximum ages and endemicity status. For 13 archipelagos, which accounted for more than a third of the total species, over 90% of the species were sampled in phylogenies.

DAISIE

We used the method DAISIE[10] to estimate rates of species accumulation (colonization, speciation and extinction) on the archipelagos. The model assumes that after the origin of an island, species can colonize from a mainland pool. Once a species has colonized, it may remain similar to its mainland ancestor (non-endemic species), become endemic through anagenetic speciation (new endemic species is formed without lineage splitting on the island), split into new species via cladogenetic speciation and/or go extinct. A carrying capacity (that is, the maximum number of species each colonist lineage can attain) is implemented, such that rates of cladogenesis and colonization decline with increasing number of species in the colonizing clade.

The only effect of anagenesis in DAISIE is that the colonizing species becomes endemic, because further anagenesis events on the endemic species do not leave a signature in the data. However, the rate of anagenesis is not systematically underestimated. Suppose the rate was higher; it would then follow that colonizing species would also become endemic faster, and we would see more endemic species. Thus, the number of endemic species determines the rate of anagenesis, and DAISIE estimates the true rate of anagenesis without systematic bias. Further anagenesis events do not have an effect on the state variables, and hence do not enter the equations anymore.

In its parameterization of extinction, DAISIE accounts for the fact that there may have been several lineages that were present on the insular system in the past but that went completely extinct due to natural causes, leaving no extant descendants. Simulations have shown that the rate of natural extinction is usually well estimated in DAISIE (see 'Measuring precision and accuracy' and a previously published study[53]). Studies on phylogenies of single clades suggest that phylogenetic data on only extant species provide less information on extinction than on speciation (or rather diversification rates[54]). However, there is information content in such data[55], especially when diversification dynamics are dependent on diversity[56]. Moreover, here we use colonization times in addition to phylogenetic branching times to estimate extinction rates, and we are estimating hyperparameters of the theoretically and empirically suggested relationship of extinction with area. Finally, we use data from many independent colonizations, which increases the power of our statistical method considerably and decreases the bias, as maximum likelihood is known to asymptotically provide unbiased estimates.

Estimating global hyperparameters

Our aim is to examine the dependencies of the parameters that govern species assembly (colonization, extinction, cladogenesis, anagenesis (CES rates) and carrying capacity) on the features of archipelagos (area and isolation). We developed a method to estimate global hyperparameters that control the relationship between two key archipelago features (area and isolation) and archipelago-specific (local) CES rates. One can estimate directly from the global dataset the shape of the relationship between isolation and colonization rate that maximizes the likelihood for the entire dataset.

Our method finds the hyperparameters that maximize the likelihood of the entire dataset, that is, the sum of the log-likelihoods for each archipelago. We tested the hypothesis that area and distance from the nearest mainland have an effect on CES rates (cladogenesis, anagenesis,

extinction and colonization). If an effect was identified, we also estimated the scaling of the effect. We developed a set of a priori models in which the CES rates are affected by archipelago features as is often assumed in the island biogeography literature (Supplementary Table 1). For the a priori models, we considered that CES rates are determined by a power function of area or distance. In the power function, $par = par_0 I^h$ where par is the CES rate (for example, local rate of colonization), par_0 is the initial value of the biogeographical rate (for example, global initial rate of colonization), I is the physical variable (area or distance) and h is the strength of the relationship. The exponent h can be negative or positive depending on the nature of the relationship. par_0 and h are the hyperparameters. If the exponent h is estimated as zero, there is no relationship between I and the parameter. By including or excluding h from the different relationships, we can compare different models with the effects switched on or off (Supplementary Table 1; for example, in model M1 all relationships are estimated, whereas in model M2 the exponent of the relationship between anagenesis and distance is fixed to zero and thus anagenesis does not vary with distance).

In addition to the a priori models, we considered a set of post hoc models with alternative shapes of relationships. We fitted two types of post hoc models: power models and sigmoid models (Supplementary Table 1). In the post hoc power models, we modelled all parameters as in the a priori models, except for cladogenesis: we allowed cladogenesis to be dependent on both area and distance. The reason for this is that we found that the predicted number of cladogenetic species under the a priori models were not as high as observed, so we examined whether including a positive effect of distance would improve the fit. We described the relationship between area, distance and cladogenesis using different functions—one model in which there is an additive effect of area and distance (M15); and three models (M16, M17 and M18) in which the effect of area and distance is interactive. In addition, we fitted a model identical to M16 but with one parameter less (M19). The reason for this was that this parameter (y) was estimated to be zero in M16.

In the post hoc sigmoid models, we allowed the relationship between distance and a given parameter to follow a sigmoid rather a power function. The rationale for this was that we wanted to investigate whether, for birds, the effect of distance on a parameter only starts to operate after a certain distance from the mainland, as below certain geographical distances archipelagos are within easy reach for many bird species by flight, so that at these distances the island behaves almost as part of the mainland from a bird's perspective. We fitted nine different sigmoid models (Supplementary Table 1), allowing cladogenesis, anagenesis and colonization to vary with distance following a sigmoid function. The sigmoid function that we used has an additional parameter in comparison to power functions.

In total, we fitted 28 candidate models (14 a priori, 14 post hoc) to the global dataset using maximum likelihood. We fitted each model using 20 initial sets of random starting parameters to reduce the risk of being trapped in local likelihood suboptima. We used the age of each archipelago (Extended Data Table 1) as the maximum age for colonization. We assumed a global mainland species pool M of 1,000 species. The product of M and the intrinsic rate of colonization (γ_0) is constant as long as M is large enough (larger than the number of island species), and thus the chosen value of M does not affect the results.

To decide which information criterion to use to select between different models, we compared the performance of the BIC and the Akaike information criterion (AIC). We simulated 1,000 datasets each with models M9 and M19 and then fitted the M9, M14, M17 and M19 models to each of these datasets using two initial sets of starting parameters for each optimization. We found that for datasets simulated using M9 an incorrect model was preferred using AIC in 10.4% of cases, but only in 0.11% of cases when using BIC. For datasets simulated using M19 an incorrect model was preferred 12.8% of cases using AIC and 11.1% of cases using BIC. We thus compared models using BIC, as this model has lower error rates.

An alternative approach to estimating hyperparameters would be to calculate CES rates and their uncertainty independently for each archipelago and to then conduct a meta-analysis of the resulting data, including archipelago area and isolation as predictors. However, errors in parameter estimates will vary, particularly because some archipelagos have small sample sizes (only a few extant colonization events, or none at all; for example, Chagos) and are thus much less informative about underlying process[53]. Thus, maximizing the likelihood of all datasets together by estimating the hyperparameters (which is precisely our aim) is preferable. For completeness, we present CES rates estimated independently for each archipelago in Supplementary Table 6, excluding archipelagos with fewer than 6 species and for which we sampled less than 60% of the species in the phylogenies. However, as argued above we do not advocate using these parameter estimates for further analyses because the number of taxa for some of these archipelagos is still low and by excluding archipelagos with fewer than six taxa, we cannot capture the lower part of the relationship between area or isolation and CES rates.

All DAISIE analyses were run using parallel computation on the high-performance computer clusters of the University of Groningen (Peregrine cluster) and the Museum für Naturkunde Berlin. The new version of the R package DAISIE is available on GitHub.

Randomization analysis

We conducted a randomization analysis to evaluate whether there is significant signal of a relationship between area and distance and local CES rates in our global dataset. We produced 1,000 datasets with the same phylogenetic data and archipelago ages as the global dataset, but randomly reshuffled archipelago area and D_m in each dataset. We then fitted the best post hoc model to each of these 1,000 randomized datasets. If the maximum-likelihood estimates of exponent hyperparameters (that is, the strength of the relationship) in the randomized datasets were non-zero, this would indicate that the method is finding evidence for a relationship even if there is none. If, on the other hand, non-zero hyperparameters are estimated in the real data but not in the randomized datasets, this would mean that there is information in the data regarding the putative relationships.

The randomization analysis showed that in global datasets with reshuffled areas and distances the exponent hyperparameters are estimated as zero in most cases, whereas in the empirical global dataset they are not (Extended Data Fig. 3).

A posteriori simulations

We simulated 1,000 phylogenetic global datasets (41 archipelagos each) with the maximum-likelihood hyperparameters of the best a priori (M14) and post hoc models (M19). We first calculated the local CES rates for each archipelago based on their area and isolation and the hyperparameters for the model, and then used these CES rates as the parameters for the simulations using the DAISIE R package. The simulated data were used to measure bias and accuracy of the method, goodness of fit and the ability of our method to recover observed island biogeographical diversity patterns (see 'Measuring precision and accuracy of method' and 'Measuring goodness of fit' sections).

Measuring precision and accuracy of method

DAISIE estimates the CES rates with high precision and little bias[10,53]. We conducted parametric bootstrap analyses to assess whether the ability to estimate hyperparameters from global datasets is also good (Extended Data Fig. 2) and to obtain confidence intervals on parameter estimates (Extended Data Table 5). We used DAISIE to estimate hyperparameters from the M14 and M19 simulated datasets (1,000 replicates each). We measured precision and accuracy by comparing the distribution of parameters estimated from the 1,000 simulated dataset with the real parameters used to simulate the same datasets. To check whether maximum-likelihood optimizations of the simulated

global datasets converge to the same point in parameter space, we first performed a test on a subset of the simulated data. We ran optimizations with 10 random sets of initial starting values for each of 10 simulated datasets. All optimizations converged to the same likelihood and a very similar hyperparameter set; therefore, we are confident that we found the global optimum for each simulated global dataset, even for models with many parameters.

Measuring goodness of fit

We measured how well the preferred models fitted the data using different approaches. First, we examined whether our models successfully reproduced the diversity patterns found on individual archipelagos. We calculated the total number of species, cladogenetic species and independent colonizations in each archipelago for each of the 1,000 simulated datasets. We then plotted these metrics versus the observed values in the empirical data (Fig. 3 and Extended Data Fig. 4). Our preferred models have a slight tendency to overpredict species richness when there are a few species and underpredict it when there are many. We do not have a clear explanation for this. This slight deviation does not seem to be due to an additional dependence on area or distance, so an explanation should be sought in other factors that we did not model. We note that the fact that all three plots show this tendency rather than only one is to be expected because the three metrics of species richness are not entirely independent, with total species richness being the sum of the other two.

Second, we examined whether the models successfully predict the empirical relationships between area, distance and diversity metrics (total species, cladogenetic species, and number of independent colonizations). We fitted generalized linear models for each diversity metric, with quasi-Poisson family errors and log area (or distance) as predictors. We then repeated this across 1,000 independent sets of simulated data for the 41 archipelagos and compared the mean of slopes and intercepts for archipelago area and archipelago isolation to the equivalent estimates for the empirical data (Fig. 4).

Third, we estimated the pseudo-R^2 of the best model (M19) as a measure of the explanatory power of the model. We simulated two independent sets of 10,000 global datasets under M19 model (set 1 and set 2). We calculated the mean total number of species, number of cladogenetic species and colonizations for each archipelago across all datasets from set 1. For each diversity metric, we calculated a pseudo-R^2 (pseudo-R^2 observed) for which the total sum of squares was obtained from the empirical data and the residual sum of squares was calculated as the difference between empirical values and expected values (that is, the simulation means). As the model is inherently stochastic, even if the model is an accurate and complete reflection of the underlying processes then the pseudo-R^2 would tend to be <1. To estimate the distribution of pseudo-R^2 expected under the model, we treated the set-2 simulations as data and estimated the pseudo-R^2 for each (pseudo-R^2 simulated). We then calculated the ratio of the pseudo-R^2-observed values over the 10,000 pseudo-R^2-simulated values. A ratio approaching 1 would indicate that the model is explaining the observed data as well as the average dataset simulated under this process (Extended Data Fig. 5).

Sensitivity to alternative divergence times and tree topologies

Despite having sampled many new individuals from islands worldwide, given the wide geographical scale of our study we still rely on sequence data for thousands of individuals submitted to GenBank over the years. Whenever multi-loci analyses including our focal taxa were available we used them; however, these are rare (Extended Data Table 4). Therefore, the majority of our phylogenies are based on a small number of genes, and most on a single gene, cytochrome *b*, which is the most widely sequenced mitochondrial marker in birds. Although some studies on island birds have shown that colonization and diversification times derived from mitochondrial trees often do not differ much from those obtained using multiple loci[57], it is possible that in some cases the scaling and topologies of the trees might have been more accurate had we used multiple loci[58]. This is particularly relevant for recent island colonists, given incomplete lineage sorting[59]. An additional shortcoming of relying on published sequence data is that many of our DNA alignments often have substantial sections with missing data (for example, because only one small section of the gene could be sequenced and was uploaded to GenBank), which has been shown to lead to biases in branch lengths and topology[60]. While future studies using phylogenomic approaches may address these issues, obtaining tissue samples for all of these taxa will remain an obstacle for a long time.

Although DAISIE does not directly use topological information (only divergence times are used), it is possible that the true topology for a clade may differ from that of the gene tree that we have estimated and this could have an effect on our results by (1) affecting colonization and branching times (addressed in the paragraph below); or (2) by altering the number of colonization events. Alternative topologies may have led to an increase or decrease in colonization events—for instance, some species that appear to have colonized an archipelago only once may have colonized multiple times and if these re-colonizations are recent they may go undetected when using one or few loci. As with any phylogenetic study, we cannot rule out this possibility, but we assume that recent re-colonization of the archipelagos in our dataset by the same taxon is rare, as these are all oceanic and isolated. For archipelago lineages with cladogenesis (26 out of 502 lineages), alternative topologies could include non-monophyly of island radiations, with the corollary being that they would be the result of multiple colonization events. However, this seems improbable for these isolated and well-studied radiations, for which morphological evidence (for example, HBW[37]) is consistent with their monophyly as supported by existing molecular data.

Regarding scaling of divergence times, we assessed how uncertainty in our estimated node ages could influence our results by running an analysis of 100 datasets. For each dataset we sampled the node ages (that is, colonization and branching times) at random from a uniform distribution centred on the posterior mean for that node in the BEAST tree and extending twice the length of the highest posterior density (HPD) interval. For example, for a node with a 95% HPD interval of 2–3 million years in our trees, the uniform distribution was set to between 1.5 and 3.5 million years. The HPD interval will capture uncertainty under the selected phylogenetic and substitution models for the loci that we used, but we conduct our sensitivity analysis over a broader interval to accommodate the potential that the selected models and gene trees are inadequate. For cases in which using this approach meant that the lower bound of the uniform distribution was less than 0, we assigned a value of 0.00001 million years to the lower bound. We fitted the 9 best models to the 100 datasets using 5 initial starting parameters for each model (total 4,500 optimizations). We found that parameter estimates across the 100 datasets did not differ strongly from those in the main dataset (Supplementary Table 7). Notably, model selection was unaffected, with the M19 model being selected for all 100 datasets. This is because a lot of the information used for model selection is coming from the other sources of information that DAISIE uses (island age, number of species and endemicity status) rather than colonization or branching times.

The maximum-likelihood parameters of the M19 model and the resulting area and isolation dependencies for datasets D1 to D6 (discussed below) are shown in Extended Data Fig. 6 and the DAISIE R objects including these alternative datasets are available in Mendeley Data (https://doi.org/10.17632/sy58zbv3s2.2).

To account for uncertainty in the rates of molecular evolution, we repeated all BEAST dating analyses for markers that were not cytochrome *b* using (1) the previously published cytochrome *b* rate[50] (dataset D1, equal to main dataset) and (2) previously estimated marker-specific rates[41], which have also been widely used in the literature

(dataset D2). Although the trees dated using the marker-specific rates provide younger ages, we found that the DAISIE results were very similar using either approach (same model preferred and similar parameters). Therefore, in the main text we only discuss the results of analyses of D1, that is, applying the cytochrome *b* rate to all markers.

For some taxa, we did not use the stem age as the estimate of colonization time, and instead used alternative nodes (see 'Colonization and branching times'). To test whether our choice of nodes affects our main conclusions, we recoded all such taxa by extracting the stem ages and used these ages as an upper bound for colonization (DAISIE MaxAge option). We fitted all 28 models to this new dataset (D3) and found that the M19 model is preferred and that the parameters and area or isolation relationships vary only slightly from those of the main analysis. We therefore conclude that our results are robust to the node selection approach.

If extinction has been high on the mainland, or if we failed to sample the closest relatives of the island taxa, this could lead to an overestimation of colonization times when using the stem age as the precise time of colonization. To investigate how this could have influenced our results, we ran analyses of datasets in which we allowed colonization to have happened at any time since the stem age (that is, the time of divergence from the nearest relative of the taxon on the mainland). For this we used the DAISIE options Endemic_MaxAge or NonEndemic_MaxAge, which integrate over all possible ages between the given maximum age and the present (or the first branching event within the archipelago for cases in which cladogenesis has occurred). We repeated this analysis coding all stem ages as maximum ages (D4), or coding only the 25% older stem ages as maximum ages (to account for the fact that older stems have the potential to have more bias) (D5). We also ran analyses on 100 datasets (D6) for which we assigned precise younger ages by randomly selecting a value between the stem age and the present (or crown age for cladogenetic groups). For all of these datasets (D4–D6), we found that the same model (M19) was preferred, but the initial values of the biogeographical rates (cladogenesis, extinction, colonization and anagenesis) were estimated to be higher than in the main dataset. Notably, the exponent hyperparameters were similar to those in the main dataset, meaning that the shape of the relationships between parameters and area or isolation is not much affected (Extended Data Fig. 6). The only exception is perhaps anagenesis, for which the relationships varied more markedly—with isolated islands achieving very high rates for this parameter—but still agreeing with our main conclusions. Anagenesis is in general the most difficult parameter to estimate[53]. Thus, our conclusions are robust to the colonization times potentially being younger than those in our main dataset.

Sensitivity to archipelago selection and isolation metrics

The results of the following sensitivity analyses are presented in Supplementary Data 3 and the DAISIE R objects that include these alternative datasets are available in Mendeley Data (https://doi.org/10.17632/sy58zbv3s2.2).

To test whether the inclusion of both true archipelagos and single islands in our dataset could affect the results, we repeated analyses excluding single island units and found that the same model was preferred. The estimated initial rate of cladogenesis (λ^c_0) is higher if we exclude single islands, but this parameter is not different from a distribution of parameters estimated from datasets generated using a stratified-random sampling of both archipelagos and single islands.

Alternative isolation metrics to D_m have been shown to explain varying and often higher amounts of variation in species richness on islands[61]. We tested two alternative metrics: distance to the nearest larger or equivalent-sized landmass (D_b), and the mean between D_m and D_b (metrics given in Supplementary Data 2). We found that the same DAISIE model with very similar parameters was preferred in both cases, and we therefore used only the D_m metric, as this is more similar to the original model of MacArthur and Wilson.

The Mascarenes (Mauritius Island, Reunion and Rodrigues) are often treated as a single biogeographical unit in analyses. We chose to analyse them as independent units because (1) the distance between islands is much greater than our threshold for archipelago definition (more than 500 km between Mauritius Island and Rodrigues; more than 170 km between Reunion and Mauritius Island); (2) only two species of our target group are shared between the islands (*Terpsiphone bourbonnensis* is found in Mauritius Island and Reunion; and *Psittacula eques* is found in Mauritius Island and extirpated from Reunion), suggesting low connectivity; (3) although there are three clades whose branching events took place within the Mascarenes (*Coracina*, *Pezophaps* and *Raphus*, and *Zosterops*), the remaining species result from independent colonizations, suggesting that the three islands behave mostly as three different biogeographical units. We nevertheless ran an analysis treating the islands as a single archipelagic unit and found that the same model was preferred and with similar parameter estimates, and we therefore discuss only the results treating them as separate.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

New sequence data produced for this study have been deposited in GenBank with the accession codes: MH307408–MH307656. The following datasets have been deposited in Mendeley: DNA alignments (https://doi.org/10.17632/vf95364vx6.1), new phylogenetic trees produced for this study (https://doi.org/10.17632/p6hm5w8s3b.2), and DAISIE R objects (https://doi.org/10.17632/sy58zbv3s2.2). The 11 previously published trees are available upon request.

Code availability

The custom computer code used for this study is freely available in the DAISIE R package (https://github.com/rsetienne/DAISIE).

31. Plummer, P. S. & Belle, E. R. Mesozoic tectono-stratigraphic evolution of the Seychelles microcontinent. *Sediment. Geol.* **96**, 73–91 (1995).
32. Weigelt, P., Jetz, W. & Kreft, H. Bioclimatic and physical characterization of the world's islands. *Proc. Natl Acad. Sci. USA* **110**, 15307–15312 (2013).
33. Norder, S. J. et al. Beyond the Last Glacial Maximum: island endemism is best explained by long-lasting archipelago configurations. *Glob. Ecol. Biogeogr.* **28**, 184–197 (2019).
34. Thomson, J. & Walton, A. Redetermination of chronology of Aldabra atoll by $^{230}Th/^{234}U$ dating. *Nature* **240**, 145–146 (1972).
35. Price, J. P. & Clague, D. A. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc. R. Soc. B* **269**, 2429–2435 (2002).
36. Valente, L. et al. Equilibrium bird species diversity in Atlantic islands. *Curr. Biol.* **27**, 1660–1666 (2017).
37. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & Kirwan, G. (eds.) *Handbook of the Birds of the World Alive* (Lynx Edicions, 2018).
38. Valente, L., Etienne, R. S. & Dávalos, L. M. Recent extinctions disturb path to equilibrium diversity in Caribbean bats. *Nat. Ecol. Evol.* **1**, 0026 (2017).
39. Steadman, D. W. *Extinction and Biogeography of Tropical Pacific Birds* (Univ. Chicago Press, 2006).
40. Cheke, A. & Hume, J. P. *Lost Land of the Dodo: The Ecological History of Mauritius, Réunion and Rodrigues* (Bloomsbury, 2010).
41. Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M. & Fleischer, R. C. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr. Biol.* **21**, 1838–1844 (2011).
42. Rando, J. C., Pieper, H., Olson, S. L., Pereira, F. & Alcover, J. A. A new extinct species of large bullfinch (Aves: Fringillidae: *Pyrrhula*) from Graciosa Island (Azores, North Atlantic Ocean). *Zootaxa* **4282**, 567–583 (2017).
43. Illera, J. C., Rando, J. C., Richardson, D. S. & Emerson, B. C. Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quat. Sci. Rev.* **50**, 14–22 (2012).
44. Hume, J. P., Martill, D. & Hing, R. A terrestrial vertebrate palaeontological review of Aldabra atoll, Aldabra group, Seychelles. *PLoS ONE* **13**, e0192675 (2018).
45. Cheke, A. S. Extinct birds of the Mascarenes and Seychelles—a review of the causes of extinction in the light of an important new publication on extinct birds. *Phelsuma* **21**, 4–19 (2013).
46. Hume, J. P. & Walters, M. *Extinct Birds* (A&C Black, 2012).
47. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

48. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
49. Posada, D. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
50. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17**, 2321–2328 (2008).
51. Field, D. J. et al. Timing the extant avian radiation: the rise of modern birds, and the importance of modeling molecular rate variation. *PeerJ Preprints* **7**, e27521v1 (2019).
52. Cicero, C. & Johnson, N. K. Higher-level phylogeny of new world Vireos (Aves: Vireonidae) based on sequences of multiple mitochondrial DNA genes. *Mol. Phylogenet. Evol.* **20**, 27–40 (2001).
53. Valente, L., Phillimore, A. B. & Etienne, R. S. Using molecular phylogenies in island biogeography: it's about time. *Ecography* **41**, 1684–1686 (2018).
54. Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824 (2010).
55. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* **344**, 305–311 (1994).
56. Etienne, R. S. et al. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* **279**, 1300–1309 (2012).
57. Stervander, M. et al. Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.) — phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Mol. Ecol.* **24**, 2477–2494 (2015).
58. Ogilvie, H. A., Heled, J., Xie, D. & Drummond, A. J. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Syst. Biol.* **65**, 381–396 (2016).
59. Maddison, W. P. & Knowles, L. L. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21–30 (2006).
60. Lemmon, A. R., Brown, J. M., Stanger-Hall, K. & Lemmon, E. M. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**, 130–145 (2009).
61. Weigelt, P. & Kreft, H. Quantifying island isolation—insights from global patterns of insular plant species richness. *Ecography* **36**, 417–429 (2013).
62. Nielson, D. L. & Sibbett, B. S. Geology of Ascension Island, South Atlantic Ocean. *Geothermics* **25**, 427–448 (1996).
63. Ramalho, R. S. et al. Emergence and evolution of Santa Maria Island (Azores)—the conundrum of uplifted islands revisited. *Geol. Soc. Am. Bull.* **129**, 372–390 (2017).
64. Hearty, P. J. & Olson, S. L. Geochronology, biostratigraphy, and changing shell morphology in the land snail subgenus *Poecilozonites* during the Quaternary of Bermuda. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **293**, 9–29 (2010).
65. Carracedo, J. C. & Troll, V. R. *The Geology of the Canary Islands* (Elsevier, 2016).
66. Ramalho, R. *Building the Cape Verde Islands* (Springer, 2011).
67. Eisenhauer, A., Heiss, G. A., Sheppard, C. R. C. & Dullo, W. C. in *Ecology of the Chagos Archipelago* (eds. Sheppard, C. R. C. & Seaward, M. R. D.) 21–31 (Linnean Society Occasional Publications, 1999).
68. Campbell, H. J. Fauna and flora of the Chatham Islands: less than 4 m.y. old. *Geological Society of New Zealand Miscellaneous Publication 97* (eds Cooper, R. A. & Jones, C.) 15–16 (Geological Society of New Zealand, 1998).
69. Bullough, F. History and geology of Christmas Island. *Geological Society of London Blog* https://blog.geolsoc.org.uk/2013/12/18/door-18-history-and-geology-of-christmas-island/ (2013).
70. Castillo, P. et al. Anomalously young volcanoes on old hot-spot traces: I. Geology and petrology of Cocos Island. *Geol. Soc. Am. Bull.* **100**, 1400–1414 (1988).
71. Woodroffe, C. D., Veeh, H. H., Falkland, A. C., McLean, R. F. & Wallensky, E. Last interglacial reef and subsidence of the Cocos (Keeling) Islands, Indian Ocean. *Mar. Geol.* **96**, 137–143 (1991).
72. Nougier, J., Cantagrel, J. M. & Karche, J. P. The Comores archipelago in the western Indian Ocean: volcanology, geochronology and geodynamic setting. *J. Afr. Earth Sci.* **5**, 135–144 (1986).
73. Almeida, F. in *Sítios Geológicos e Paleontológicos do Brasil* (eds Schobbenhaus, C. et al.) 361–368 (Comissão Brasileira de Sítios Geológicos e Paleobiológicos, 2000).
74. Ali, J. R. & Aitchison, J. C. Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J. Biogeogr.* **41**, 1227–1241 (2014).
75. Ryan, P. G. in *Encyclopedia of Islands* (eds. Gillespie, R. & Clague, D.) 929–932 (Univ. California Press, 2009).
76. Batiza, R. Petrology and chemistry of Guadalupe Island: an alkalic seamount on a fossil ridge crest. *Geology* **5**, 760–764 (1977).
77. Stuessy, T. F., Foland, K. A., Sutter, J. F., Sanders, R. W. & Silva O., M. Botanical and geological significance of potassium–argon dates from the Juan Fernandez islands. *Science* **225**, 49–51 (1984).
78. McDougall, I., Embleton, B. J. J. & Stone, D. B. Origin and evolution of Lord Howe Island, Southwest Pacific Ocean. *J. Geol. Soc. Aust.* **28**, 155–176 (1981).
79. Mata, J. et al. in *Geologia de Portugal, Volume II Geologia Meso-cenozoica de Portugal* (eds. Dias, R. et al.) 691–746 (Escolar Editora, 2013).
80. Guille, G. et al. Les marquises (Polynésie Françaises): un archipel intraocéanique atypique. *Geol. France* **2**, 5–36 (2002).
81. Montaggioni, L. & Nativel, P. *La Reunion, Ile Maurice. Géologie et Aperçus Biologiques, Plantes et Animaux* (Masson, 1988).
82. Grandcolas, P. et al. New Caledonia: a very old Darwinian island? *Phil. Trans. R. Soc. B* **363**, 3309–3317 (2008).
83. Anthoni, J. Geography and geology of Niue. http://www.seafriends.org.nz/niue/geo.htm (2005).
84. Jones, J. G. & McDougall, I. Geological history of Norfolk and Philip islands, southwest Pacific Ocean. *J. Geol. Soc. Aust.* **20**, 239–254 (1973).
85. Suzuki, M., Taisuke, S. & Hideo, T. *Nomination of the Ogasawara Islands for Inscription on the World Heritage List* (Government of Japan, 2010).
86. Neall, V. E. & Trewick, S. A. The age and origin of the Pacific islands: a geological overview. *Phil. Trans. R. Soc. B* **363**, 3293–3308 (2008).
87. Hekinian, R. et al. The Pitcairn hotspot in the South Pacific: distribution and composition of submarine volcanic sequences. *J. Volcanol. Geotherm. Res.* **121**, 219–245 (2003).
88. Vezzoli, L. & Acocella, V. Easter Island, SE Pacific: an end-member type of hotspot volcanism. *Bull. Geol. Soc. Am.* **121**, 869–886 (2009).
89. Gillot, P.-Y., Lefèvre, J.-C. & Nativel, P.-E. Model for the structural evolution of the volcanoes of Réunion Island. *Earth Planet. Sci. Lett.* **122**, 291–302 (1994).
90. Safford, R. & Hawkins, F. *The Birds of Africa: Volume VIII: The Malagasy Region: Madagascar, Seychelles, Comoros, Mascarenes* (A&C Black, 2013).
91. Baker, I., Gale, N. H. & Simons, J. Geochronology of the St Helena volcanoes. *Nature* **215**, 1451–1456 (1967).
92. Duncan, R. A. in *Investigations of the Northern Melanesian Borderland, Earth Science Series* (Circum Pacific Council Publications, 1985).
93. Lee, D. C., Halliday, A. N., Fitton, J. G. & Poli, G. Isotopic variations with distance and time in the volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth Planet. Sci. Lett.* **123**, 119–138 (1994).
94. Geldmacher, J., Hoernle, K., Van Den Bogaard, P., Zankl, G. & Garbe-Schönberg, D. Earlier history of the ≥70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen Archipelago and neighboring seamonts in the Eastern North Atlantic. *J. Volcanol. Geotherm. Res.* **111**, 55–87 (2001).
95. Clouard, V. & Bonneville, A. A in *Plates, Plumes and Paradigms* (eds Foulger, G. R. et al.) 71–90 (Geological Society of America, 2005).
96. Bohrson, W. A. et al. Prolonged history of silicic peralkaline volcanism in the eastern Pacific Ocean. *J. Geophys. Res.* **101**, 11457–11474 (1996).
97. Kroenke, L. W. in *The Origin and Evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: Patterns and Processes* (eds. Keast, A. & Miller, S.) 19–34 (SPD Academic Publishing, 1996).
98. Ollier, C. D. Geomorphology of South Atlantic volcanic islands. Part I: the Tristan da Cunha group. *Z. Geomorphol.* **28**, 367–382 (1984).
99. Kocher, T. D. et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl Acad. Sci. USA* **86**, 6196–6200 (1989).
100. Dietzen, C., Witt, H.-H. & Wink, M. The phylogeographic differentiation of the European robin *Erithacus rubecula* on the Canary Islands revealed by mitochondrial DNA sequence data and morphometrics: evidence for a new robin taxon on Gran Canaria? *Avian Sci.* **3**, 115–132 (2003).
101. Edwards, S. V., Arctander, P. & Wilson, A. C. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B* **243**, 99–107 (1991).
102. Helm-Bychowski, K. & Cracraft, J. Recovering phylogenetic signal from DNA sequences: relationships within the corvine assemblage (class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-*b* gene. *Mol. Biol. Evol.* **10**, 1196–1214 (1993).
103. Warren, B. H., Bermingham, E., Bowie, R. C. K., Prys-Jones, R. P. & Thébaud, C. Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (*Nectarinia*: Nectariniidae). *Mol. Phylogenet. Evol.* **29**, 67–85 (2003).
104. Farrington, H. L., Lawson, L. P., Clark, C. M. & Petren, K. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution* **68**, 2932–2944 (2014).
105. Warren, B. H. et al. Hybridization and barriers to gene flow in an island bird radiation. *Evolution* **66**, 1490–1505 (2012).
106. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thebaud, C. Tracking island colonization history and phenotypic shifts in Indian Ocean bulbuls (*Hypsipetes*: Pycnonotidae). *Biol. J. Linn. Soc.* **85**, 271–287 (2005).
107. Andersen, M. J., Hosner, P. A., Filardi, C. E. & Moyle, R. G. Phylogeny of the monarch flycatchers reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation. *Mol. Phylogenet. Evol.* **83**, 118–136 (2015).
108. Sari, E. H. R. & Parker, P. G. Understanding the colonization history of the Galápagos flycatcher (*Myiarchus magnirostris*). *Mol. Phylogenet. Evol.* **63**, 244–254 (2012).
109. Chaves, J. A., Parker, P. G. & Smith, T. B. Origin and population history of a recent colonizer, the yellow warbler in Galápagos and Cocos Islands. *J. Evol. Biol.* **25**, 509–521 (2012).
110. Martínez-Gómez, J. E., Barber, B. R. & Peterson, A. T. Phylogenetic position and generic placement of the Socorro wren (*Thryomanes sissonii*). *Auk* **122**, 50–56 (2005).
111. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thébaud, C. Immigration, species radiation and extinction in a highly diverse songbird lineage: white-eyes on Indian Ocean islands. *Mol. Ecol.* **15**, 3769–3786 (2006).
112. McGuire, J. A. et al. Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* **24**, 910–916 (2014).
113. Derryberry, E. P. et al. Lineage diversification and morphological evolution in a large-scale continental radiation: the neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution* **65**, 2973–2986 (2011).
114. Jønsson, K. A. et al. A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). *Mol. Phylogenet. Evol.* **94**, 87–94 (2016).
115. Scofield, R. P. et al. The origin and phylogenetic relationships of the New Zealand ravens. *Mol. Phylogenet. Evol.* **106**, 136–143 (2017).
116. Cibois, A., Thibault, J. C., Bonillo, C., Filardi, C. E. & Pasquet, E. Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Mol. Phylogenet. Evol.* **110**, 19–26 (2017).
117. Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G. & Milá, B. Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: phylogeographic and phylogenomic evidence. *Mol. Ecol.* **25**, 6175–6195 (2016).
118. Marki, P. Z. et al. Supermatrix phylogeny and biogeography of the Australasian Meliphagides radiation (Aves: Passeriformes). *Mol. Phylogenet. Evol.* **107**, 516–529 (2017).
119. Fuchs, J. et al. Long-distance dispersal and inter-island colonization across the western Malagasy Region explain diversification in brush-warblers (Passeriformes: *Nesillas*). *Biol. J. Linn. Soc.* **119**, 873–889 (2016).

120. Cibois, A. et al. Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Mol. Phylogenet. Evol.* **70**, 442–453 (2014).
121. Carmi, O., Witt, C. C., Jaramillo, A. & Dumbacher, J. P. Phylogeography of the vermilion flycatcher species complex: multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Mol. Phylogenet. Evol.* **102**, 152–173 (2016).
122. Cornetti, L. et al. The genome of the 'great speciator' provides insights into bird diversification. *Genome Biol. Evol.* **7**, 2680–2691 (2015).

Acknowledgements We thank the skilled guides and field assistants who helped with sample collection in the field; the ornithologists and collection curators who were kind enough to reply to requests for material; T. von Rintelen, K. von Rintelen and C. Zorn for support or advice; A. Pigot for comments on the manuscript; N. Bunbury (Seychelles Islands Foundation), who organized sample loans of Aldabra island; J. van de Crommenacke, J. Groombridge and H. Jackson for providing samples or DNA sequences; J. A. Alcover, J. C. Rando, F. Sayol and S. Faurby for sharing data on extinct species; C. Baeta, M. Hammers, J. Hume, D. Shapiro, J. Varela and P. Cascão for permission to use photographs or illustrations; P. Hearty, R. Stern and M. Reagan for expertise on island geological ages; A. Cibois, J. McGuire, H. Lerner, P. Marki, B. Milá, G. Friis, J. Fuchs, J. P. Dumbacher and O. Carmi for providing phylogenetic data; P. Weigelt for map data. We thank the following for permission to obtain new samples or to access existing samples, and for logistic support (locations are given in brackets): A. Carvalho and the Department of the Environment (São Tomé and Príncipe); J. Obiang, N. Calvo and the Universidad Nacional de Guinea Ecuatorial for Bioko and Annobón samples (Equatorial Guinea); the Ministry of Environment, Energy and Climate Change of the Republic of Seychelles, the Seychelles Bureau of Standards, BirdLife Seychelles and Seychelles Islands Foundation (Seychelles); Centre National de Documentation et de Recherche Scientifique (Grande Comore & Anjouan), Action Comores, Direction de l'Agriculture et de la Foret (Mayotte) (Comoros); Ministere des Eaux et Forets (Madagascar) and the Madagascar Institute pour la Conservation des Ecosystemes Tropicaux (Madagascar); Mauritius National Parks and Conservation Service and Mauritius Wildlife Foundation (Mauritius); O. Hébert, W. Waheoneme, N. Clark, the Direction de L'Environment (South Province), Direction du Développement Economique (Loyalty Islands Province), and local chiefs and landowners (New Caledonia); Moroccan Environment Ministry (Morocco); Cape Verde Agriculture and Environment Ministry (Cape Verde); F. Njie and the Limbe Botanical and Zoological Garden (Cameroon); Station de Recherche de l'IRET at Ipassa-Makokou (Gabon); Fernanda Lages (ISCED-Huíla) (Angola); the regional governments of Andalucía and the Canary Islands (Spain); regional governments of Madeira and the Azores (Portugal). We thank the Department of Ornithology and Mammalogy of the California Academy of Sciences (L. Wilkinson and M. Flannery) for loaning Galápagos samples; the Natural History Museum at Tring (M. Adams) for loaning Comoros samples; the Stuttgart State Museum of Natural History for loaning stonechat samples from Madagascar. S. Block assisted with cluster analyses at the Museum für Naturkunde. The Center for Information Technology of the University of Groningen provided support and access to the Peregrine high-performance computing cluster. L.V. was funded by the German Science Foundation (DFG Research grant VA 1102/1-1), the Alexander von Humboldt Foundation, the Brandenburg Postdoc Prize 2015 and by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO); R.S.E. was supported by a NWO VICI grant; M.M. was supported by the Portuguese Science and Technology Foundation (post-doctoral grant: SFRH/BPD/100614/2014); S.M.C. was supported by the National Geographic Society (CRE grant 9383-13); J.C.I. was supported by the Spanish Ministry of Science, Innovation and Universities (PGC2018-097575-B-I00) and by a GRUPIN research grant from the Regional Government of Asturias (IDI/2018/000151); and C.T. was supported by the 'Laboratoire d'Excellence' TULIP (ANR-10-LABX-41).

Author contributions L.V., A.B.P. and R.S.E. designed the study, developed the analytical framework and performed statistical analyses. L.V. compiled the data, conducted most of the analyses and wrote the first draft. R.S.E. developed the likelihood method. A.B.P. and R.S.E. contributed substantially to the writing. M.M., B.H.W., S.M.C., J.C.I. and C.T. provided expertise on island birds, collected bird tissue samples, and provided molecular and/or phylogenetic data. K.H. and J.C.I. performed laboratory work. R.T. contributed to molecular analyses. T.A. performed analyses. All authors commented on the draft.

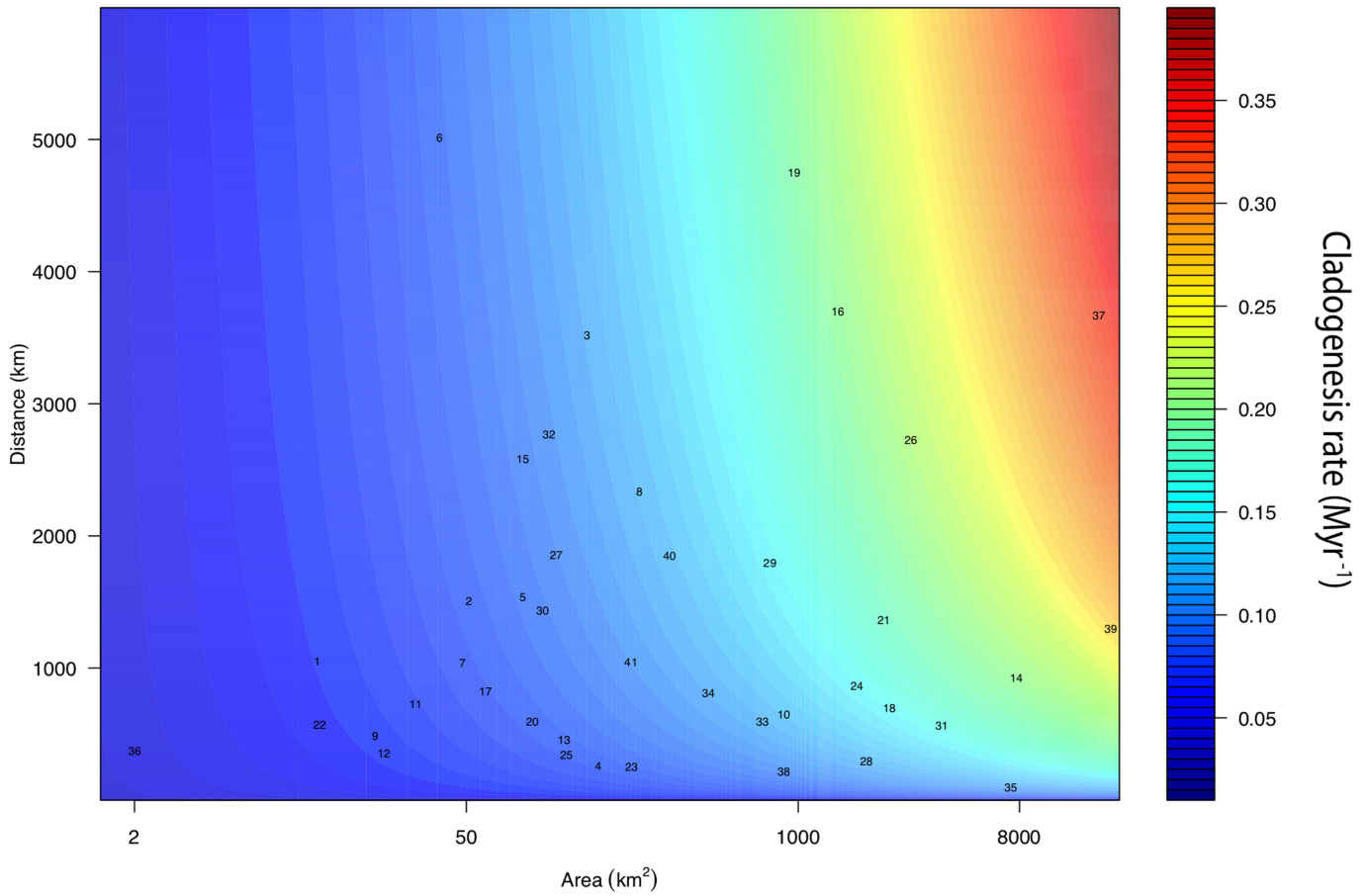
Competing interests The authors declare no competing interests.

Additional information

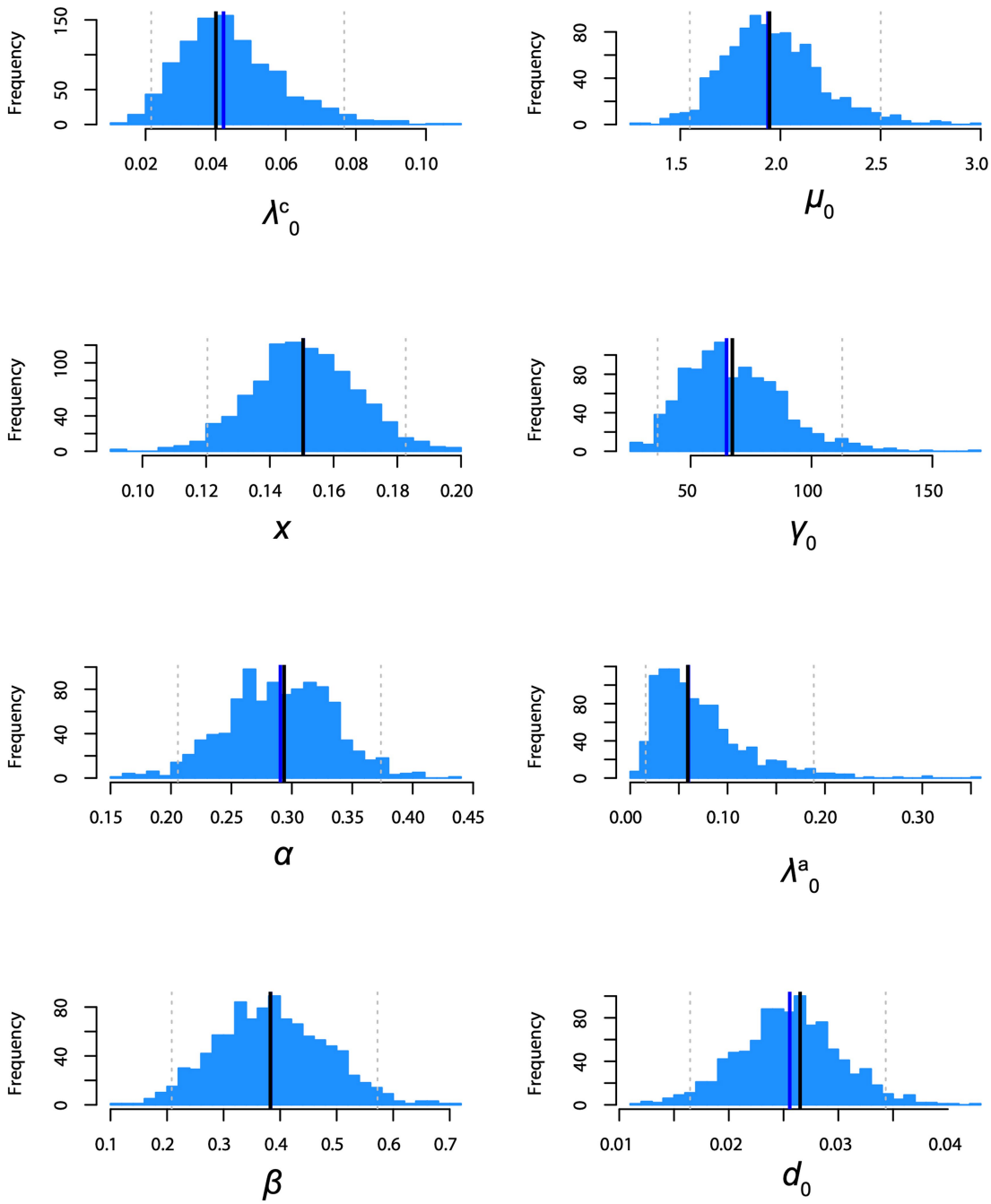
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2022-5.

Correspondence and requests for materials should be addressed to L.V.

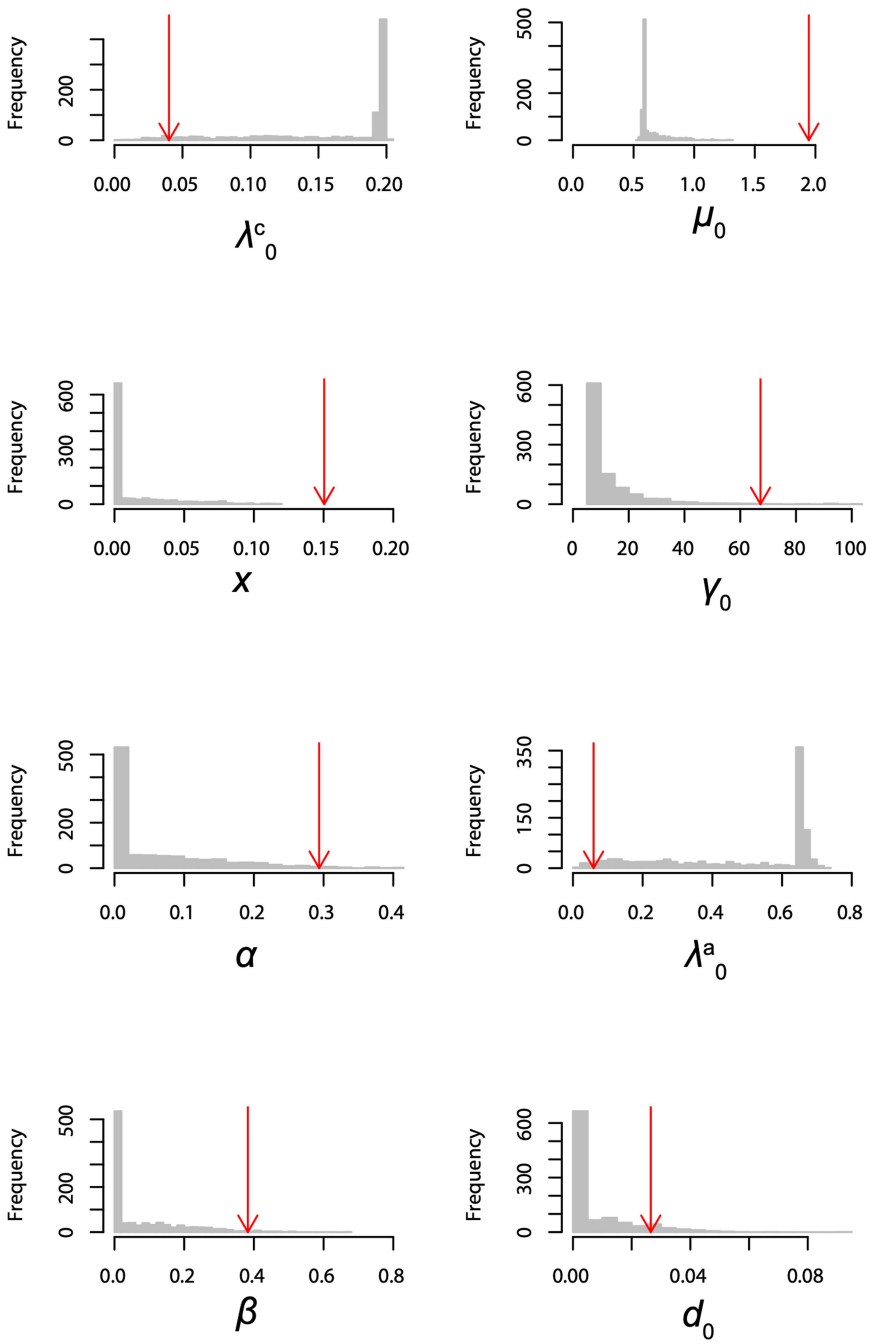
Peer review information *Nature* thanks Thomas Matthews, Kostas Triantis, Jason Weir and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



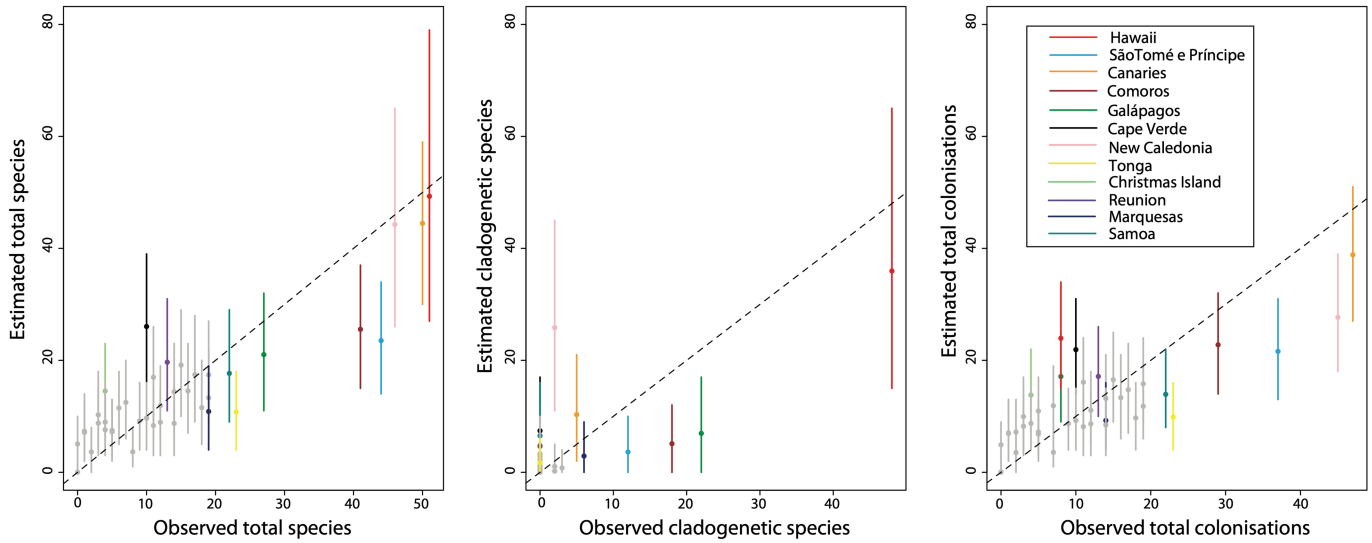
Extended Data Fig. 1 | Variation of cladogenesis with isolation and area. Contour plot showing how the local rate of cladogenesis varies with area and D_m assuming the maximum-likelihood global hyperparameters of the M19 model (equations describing the relationships are provided in Supplementary Table 1). Numbers correspond to the archipelago numbers from Fig. 1 and show the local cladogenesis rates for each of the archipelagos in our dataset. Area is shown as a log scale.



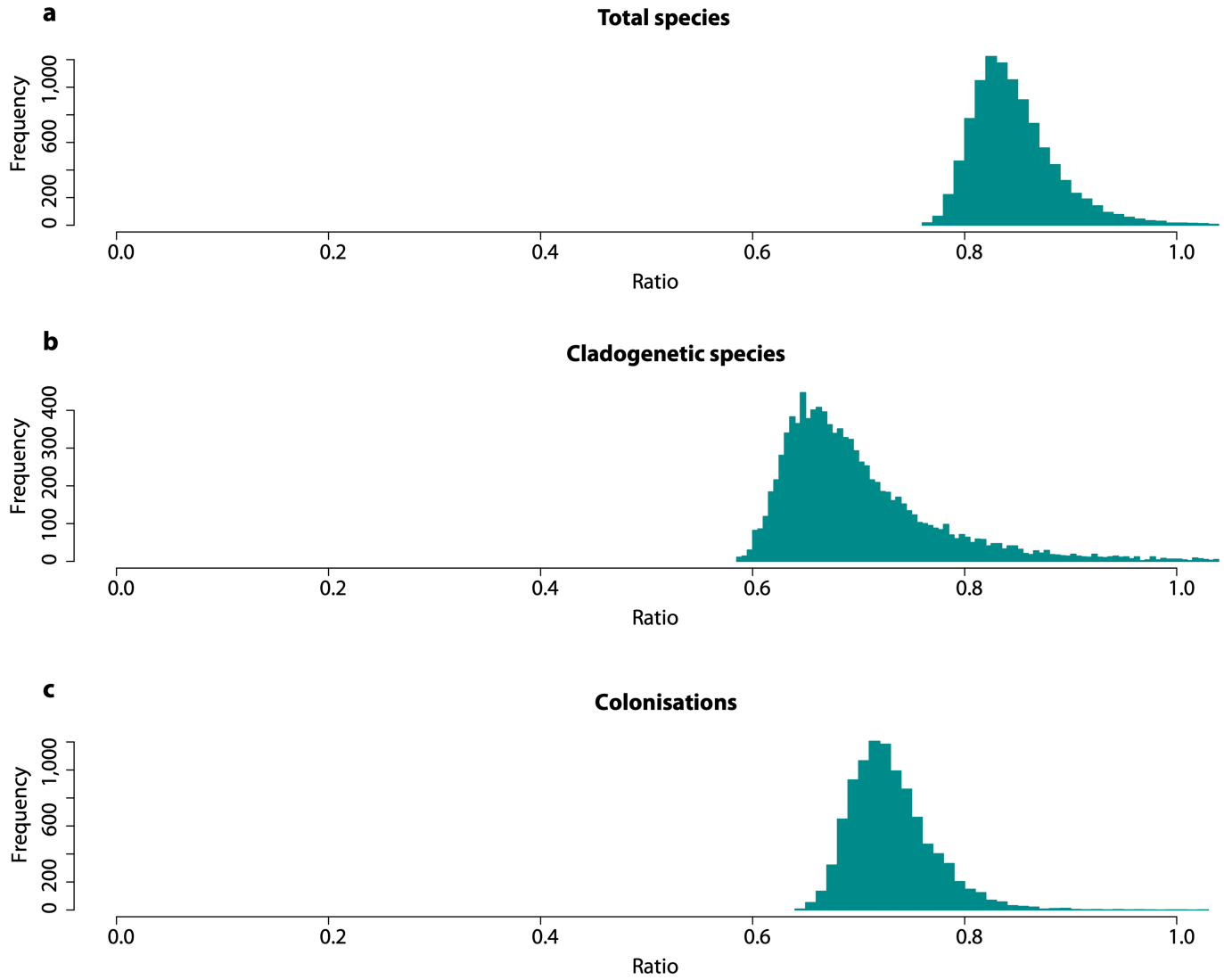
Extended Data Fig. 2 | Bootstrap precision estimates of the parameters of the M19 model. Parametric bootstrap analysis fitting the M19 model to 1,000 global datasets simulated with maximum-likelihood parameters of the M19 model. Plots are frequency histograms of estimated parameters. Black lines show the median estimated values across all simulations and the blue lines the simulated values. Dashed lines show 2.5–97.5 percentiles. Parameters are explained in Supplementary Table 1. Bootstrap parameter estimates for the M14 model are shown in Extended Data Table 5.



Extended Data Fig. 3 | Randomization analysis of the M19 model. Distribution of global hyperparameters estimated from each of 1,000 datasets with the same phylogenetic data as our main global dataset but randomly reshuffling archipelago area and isolation among the 41 archipelagos in the dataset. Grey histograms show DAISIE maximum-likelihood parameter estimates for the M19 model. Red arrows show the estimated parameter from the real data. In most cases, the hyperparameters describing the exponent of the power models (x, α, β and d_0) are estimated as zero in the reshuffled datasets, which is not the case in the real data (red). Parameters are explained in Supplementary Table 1.



Extended Data Fig. 4 | Goodness of fit of the preferred model (M19). Plots show the observed total number of species, cladogenetic species and colonizations versus those simulated by the model. Median and 95% percentiles are shown for 1,000 simulations of each archipelago. Selected archipelagos mentioned in the main text or well-known archipelagos for which one or more of the diversity metrics are under- or overestimated are highlighted in colour. Dashed line is $y = x$. See also Fig. 3.

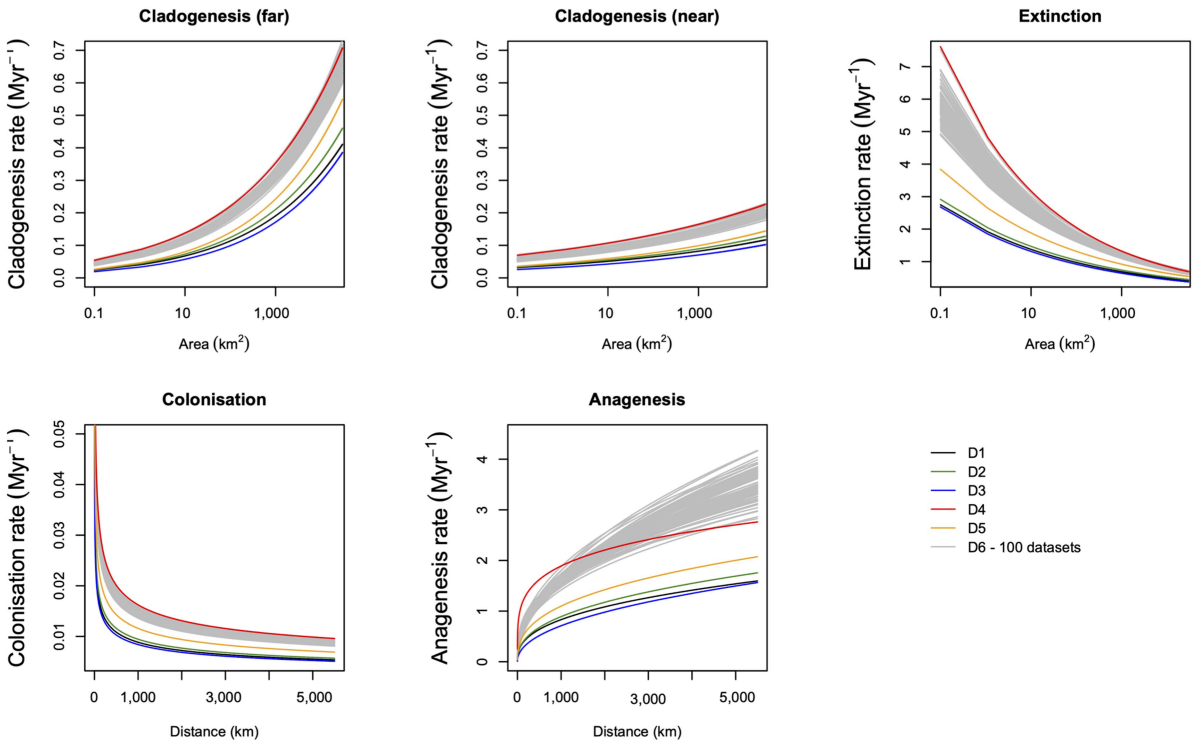


Extended Data Fig. 5 | Ratio of pseudo-R^2 observed over pseudo-R^2 simulated. Estimates were based on 10,000 datasets simulated using the M19 model. A ratio centred on 1 would indicate that the model explains the observed data as well as it is able to explain the average dataset simulated under the maximum-likelihood parameters.

a

		Cladogenesis			Extinction		Colonisation		Anagenesis	
Dataset	**Characteristics**	λ^c_0	**y**	***d*0**	μ_0	**x**	γ_0	α	λ^a_0	β
D1	Main dataset. cyt-*b* molecular rate applied to all markers.	0.040	0	0.027	1.946	0.150	67.256	0.294	0.059	0.383
D2	Molecular rates from Lerner et al. applied to non-cyt-*b* markers	0.043	0	0.027	2.070	0.149	73.460	0.297	0.057	0.398
D3	Same as D1 but choosing stem ages for all colonisation times	0.033	0	0.028	1.877	0.155	63.106	0.292	0.029	0.463
D4	Same as D1 but using all ages as a maximum age	0.086	0	0.024	4.914	0.190	137.196	0.309	0.408	0.222
D5	Same as D1 but considering the 25% oldest stem ages in the dataset as maximum ages	0.046	0	0.028	2.687	0.155	94.449	0.304	0.083	0.373
D6	Same as D1 but with all ages younger (100 datasets with random younger ages)	0.06 - 0.08	0	0.02 - 0.03	3.47 - 4.48	0.16 - 0.19	113.15 -132	0.30 - 0.31	0.04 - 0.12	0.38 - 0.52

b



Extended Data Fig. 6 | Sensitivity to colonization and branching times. **a**, Maximum-likelihood parameter estimates of the M19 model (preferred model) for datasets differing in colonization and branching times. D6 represents 100 datasets, therefore, the 2.5 and 97.5 percentiles are shown. Parameter symbols are described in Supplementary Table 1. **b**, Estimated relationships between island area and isolation and local island biogeography parameters for each dataset. Under the M19 model, cladogenesis rate increases with both area and isolation, and thus plots for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown.

Extended Data Table 1 | Archipelago characteristics and references for island geological ages

Archipelago	Area (km^2)	Distance nearest mainland (km)	Age (Ma)	Reference age	Number of species in focal group	Species sampled in phylogenies (percentage)	Colonisations sampled in phylogenies (percentage)
Aldabra Group	180	261	1	*	12	11 (91.67)	11 (91.67)
Ascension	91	1537	1	[62]	0	NA (NA)	NA (NA)
Azores	2387	1365	6.3	[63]	17	13 (76.47)	13 (76.47)
Bermuda	53	1040	2	[64]†	5	2 (40)	2 (40)
Canary Islands	7493	96	21	[65]	50	44 (88)	41 (87.23)
Cape Verde	4033	570	15.8	[66]	10	10 (100)	10 (100)
Chagos	56.13	1510	0.0065	[67]	0	NA (NA)	NA (NA)
Chatham	966	650	3	[68]	14	13 (92.86)	13 (92.86)
Christmas Island	135	345	10	[69]	4	2 (50)	2 (50)
Cocos (Costa Rica)	24	491	2.4	[70]	4	2 (50)	2 (50)
Cocos (Keeling)	14.2	1054	0.003	[71]	0	NA (NA)	NA (NA)
Comoros	2033	297	15	[72]	41	40 (97.56)	28 (96.55)
Fernando de Noronha	26	360	3.3	[73]	3	1 (33.33)	1 (33.33)
Galápagos	7880	928	4	[74]	27	26 (96.3)	8 (100)
Gough	91	2582	4	[75]	1	1 (100)	1 (100)
Guadalupe	244	254	7	[76]	11	10 (90.91)	10 (90.91)
Hawaii	16624.6	3670	29.8	[35]	51	31 (60.78)	6 (75)
Juan Fernández	99.67	600	5.8	[77]	6	6 (100)	5 (100)
Lord Howe	14.55	571	6.9	[78]	11	9 (81.82)	9 (81.82)
Madeira	798	600	18.8	[79]	19	17 (89.47)	17 (89.47)
Marianas (with Guam)	852	1800	15	‡	19	11 (57.89)	11 (57.89)
Marquesas	1063	4750	5.5	[80]	19	12 (63.16)	7 (50)
Mauritius Isl.	1865	867	8.9	[81]	15	9 (60)	9 (60)
New Caledonia	18576	1300	37	[82]	46	38 (82.61)	37 (82.22)
Niue	261.46	2340	2	[83]	5	3 (60)	3 (60)
Norfolk	34.6	730	3.05	[84]	14	11 (78.57)	11 (78.57)
Ogasawara	65	827	5	[85]	10	5 (50)	5 (50)
Palau	488	815	20.1	[86]	16	10 (62.5)	10 (62.5)
Pitcairn	42.8	5015	1.1	[87]	8	3 (37.5)	2 (28.57)
Rapa Nui	163.6	3519	0.78	[88]	2	0 (0)	0 (0)
Reunion	2512	700	5	[89]	13	9 (69.23)	9 (69.23)
Rodrigues	109	1440	15	[90]	9	4 (44.44)	4 (44.44)
Saint Helena	123.28	1856	14.5	[91]	3	0 (0)	0 (0)
Samoa	3041	2730	13.5	[92]	22	16 (72.73)	16 (72.73)
São Tomé and Príncipe	964	219	30	[93]	44	44 (100)	37 (100)
Selvagens	2.73	373	29	[94]	1	1 (100)	1 (100)
Seychelles Inner	242.68	1048	64	[31]	12	11 (91.67)	11 (91.67)
Society Islands	1577.8	3700	4.3	[95]	18	7 (38.89)	7 (38.89)
Socorro	132	457	3.5	[96]	7	7 (100)	7 (100)
Tonga	344.4	1850	41	[97]	23	10 (43.48)	10 (43.48)
Tristan da Cunha	115.4	2770	18	[98]	4	4 (100)	2 (100)

Island ages are from previously published studies[31,35,62–98].

More data are provided in Supplementary Data 2. For archipelagos closer to Madagascar, New Guinea or New Zealand than to the continent, we use those islands as the mainland.

*A previous study[34] proposed an age of 0.125 million years, but we used an older age (see Methods).

†At least 2 million years (P. Hearty, personal communication).

‡R. Stern and M. K. Reagan, personal communication.

Extended Data Table 2 | Primer sequences used in this study

Primer name	Sequence	Reference
New primers designed		
L-cytB_bird_0	TCAACRACTCCCTAATYGACCT	This paper
H-cytB_bird_2	AGRAYTACTCCTGTGTTTCARGTYTC	This paper
L-cytB_bird_2	GARACYTGAAACACAGGAGTARTYCT	This paper
H-cytB_bird_3	TAGKGGGTTGTTTGAGCCTGWTTCGTG	This paper
L-cytB_bird_3	CACGAAWCAGGCTCAAACAACCC	This paper
H-cytB_bird_4	GGAGTAGTADGGGTGAAATGGRATTTT	This paper
H-cytB_bird_5	GGGTGTTCTACTGGTTGGCTKCC	This paper
L-cytB_bird_5	CCMCTCTCACAAAYCCTATTCTGA	This paper
L-cytB_human	TGAAACTTCGGATCCCTACTA	This paper
Published primers		
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	99
L14995	GCCCCATCCAACATCTCAGCATGATGAAACTTCCG	100
L15308	GGC TAT GTC CTC CCA TGA GGC CAA AT	101
H15767	ATGAAGGGATGTTCTACTGGTTG	101
H15917	TAGTTGGCCAATGATGATGAATGGGTGTTCTACTGGTT	100
H16065	GAGTCTTCAGTCTCTGGTTTACAAGAC	102
L-cytB_Passer	CACAGGCCTAATTAAAGCCTACCT	36
H-cytB_Passer	TTGARAATGCCAGCTTTGGGAG	36
L-cytB-Mot	CCAAATYGTTACAGGMCTCCTG	36
H-cytB-Mot	GGTGAATGAGGCTAGTTGCCCA	36

Primer sequences were designed for this study or are from previously published studies[36,99–102].

Extended Data Table 3 | The 80 alignments used in the phylogenetic analyses

Taxonomic group	Molecular marker(s)	Main source of sequences
Acrocephalus	cyt-*b*	-
Alaudidae (family)	cyt-*b*	-
Alopecoenas / *Gallicolumba*	ND2	-
Anairetes	cyt-*b*	-
Anthus	cyt-*b*	-
Aphrastura	COI	-
Bucanetes	cyt-*b*	-
Buntings *Nesospiza* / *Rowettia*	cyt-*b*	-
Carduelis	cyt-*b*	-
Chasiempis / *Elepaio*	ND2	-
Chaunoproctus	ND2	-
Cinnyris notata	ATP6	[103]
Cisticola	cyt-*b*	-
Clytorhynchus	cyt-*b*	-
Coccyzus	cyt-*b*	-
Colaptes	cyt-*b*	-
Columbiformes (order)	cyt-*b*	-
Copsychus	cyt-*b*	-
Corvus	cyt-*b*	-
Crithagra / *Serinus*	cyt-*b*	-
Cuculiformes (order)	cyt-*b*	-
Cyanistes	cyt-*b*	-
Cyanolanius	cyt-*b*	-
Dendrocopos	cyt-*b*	-
Dicrurus	cyt-*b*	-
Dumetella	ND2	-
Emberiza	cyt-*b*	-
Erithacus	cyt-*b*	-
Estrilda / *Erythrura*	cyt-*b*	-
Finches, Galápagos Cocos	cyt-*b* + Multiple	[104]
Foudia	ATP8 + ATP6 + ND3	[105]
Fregilupus	ND2	-
Fringilla	cyt-*b*	-
Haemorhous	cyt-*b*	-
Hawaiian Honeycreepers	cyt-*b*	[41]
Horornis	cyt-*b*	-
Humblotia	cyt-*b*	-
Hypsipetes	ND3	[106]
Lamprotornis	ND2	-
Lanius	cyt-*b*	-
Leptosomus	cyt-*b*	-
Lonchura	cyt-*b*	-
Loxia	cyt-*b*	-
Microeca / *Eopsaltria*	cyt-*b*	-
Mimus	ND2	-
Moho	cyt-*b*	-
Monarcha	cyt-*b*	-
Motacilla	cyt-*b*	-
Myadestes	cyt-*b*	-
Myiagra	cyt-*b* + ND2	[107]
Myiarchus	cyt-*b* + ND2	[108]
Nigrita	cyt-*b*	-
Onychognathus	ND2	-
Passer / *Petronia*	cyt-*b*	-
Petroica	cyt-*b*	-
Phylloscopus	cyt-*b*	-
Pipilo	cyt-*b*	-
Pomarea	cyt-*b*	-
Prinia	cyt-*b*	-
Progne	cyt-*b*	-
Psittaciformes	cyt-*b*	-
Pyrrhocorax	cyt-*b*	-
Pyrrhula	cyt-*b*	-
Regulus	cyt-*b*	-
Saxicola	cyt-*b*	-
Sephanoides	cyt-*b*	-
Setophaga	cyt-*b*	-
Setophaga petechia	ND2 + ATP8 + CR	[109]
Sitta	cyt-*b*	-
Strepera	cyt-*b*	-
Sturnus	ND2	-
Sunbirds (family)	cyt-*b*	-
Sylvia	cyt-*b*	-
Terpsiphone	cyt-*b*	-
Troglodytes / *Thryomanes*	ND2	[110]
Turdus	cyt-*b*	-
Upupa	cyt-*b*	-
Vidua	cyt-*b*	-
Weavers (family)	cyt-*b*	-
Zosterops (Indian, Atlantic)	cyt-*b*, ND3	[111]

Sequences were obtained from previous studies as indicated. Main source of sequences is GenBank or the new sequences produced for this study, except for the cases noted in the table, for which a matrix was directly obtained from a specific study[41,103–111]. Details on molecular rates and molecular models applied to each alignment are provided in Supplementary Table 4.

Extended Data Table 4 | Previously published dated trees used

Taxonomic group	Source	Molecular markers	Calibration method
Calypte	112	Multiple	Molecular rate
Cinclodes	113	3 mtDNA and 3 nuclear	Biogeographical
Corvides	114	Multiple	Fossils
Corvus moriorum	115	Mitogenome	Fossils
Ducula	116	ND2, COI, ND3, nuclear	Secondary & Biogeography
Junco	117	ND2 + CR + COI + ATP + nuclear	Molecular rate
Meliphagides (infraorder)	118	mtDNA and nuclear	Fossils & Secondary
Nesillas	119	ND2	Molecular rate
Ptilinopus	120	ND2, COI, ND3, nuclear	Secondary & Biogeography
Pyrocephalus	121	cyt-*b*, ND2, nuclear	Molecular rate
Zosterops (Pacific)	122	cyt-*b*, ND2, ND3, ATPase	Molecular rate

Data are from previously published studies[112–122].

Extended Data Table 5 | Bootstrap of M14 and M19 models

	Cladogenesis		Extinction		Colonisation		Anagenesis	
Model	λ^c_0	y	μ_0	x	γ_0	α	λ^a_0	β
M14	0.023 (0.01 - 0.07)	0.26 (0.13 - 0.37)	1.88 (1.48 - 2.45)	0.15 (0.11 - 0.18)	51.30 (28.86 - 89.64)	0.25 (0.17 - 0.34)	0.05 (0.01 - 0.16)	0.42 (0.24 - 0.61)
	λ^c_0	d_0	μ_0	x	γ_0	α	λ^a_0	β
M19	0.04 (0.022 - 0.077)	0.027 (0.016 - 0.034)	1.95 (1.55 - 2.50)	0.15 (0.12 - 0.18)	67.26 (36.35 - 112.71)	0.29 (0.21 - 0.37)	0.059 (0.02 - 0.19)	0.38 (0.21 -0.57)

Maximum-likelihood estimates and 95% confidence intervals of the parameters of the two best models. Confidence intervals were obtained from the bootstrap analyses. Parameter symbols are explained in Supplementary Table 1.

nature research

Corresponding author(s): Luis Valente

Last updated by author(s): Dec 6, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Google Earth Pro v7.3.2.5776; Geneious v11; Excel v16.28.
Data analysis	We added the code used for the DAISIE analyses to a new version of the DAISIE R package, which we deposited on GitHub (https://github.com/rsetienne/DAISIE). Other software used: BEAST v2.4.8; jModeltest v2.1.5; R v3.5.1; RStudio v1.1.453; BBEdit v12.1.3; Cyberduck v7.1.2; Figtree v1.4.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All underlying data are available in the manuscript as supplementary data or on online databases.
New sequence data has been uploaded to GenBank with accession numbers MH307408-MH307656.

Other data types have been uploaded to Mendeley:
DNA alignments: https://doi.org/10.17632/vf95364vx6.1
New phylogenetic trees produced for this study: https://doi.org/10.17632/p6hm5w8s3b.2
DAISIE R objects: https://doi.org/10.17632/sy58zbv3s2.2

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We produced phylogenies for island birds and developed a new method for estimating rates of speciation, colonisation and extinction from these islands and to relate them to island area and isolation on a global scale. A total of 596 bird taxa were included in the DAISIE analyses (including those taxa for which no phylogenetic data was available but which were present on the islands). The number of individuals sampled per taxon included in the phylogenetic analyses varied between 1 and 15.
Research sample	We did not conduct experiments. Our samples are bird specimens whose DNA was used for phylogenetic analyses. Our sampling focused on native resident terrestrial birds from 41 archipelagos (listed in Figure 1) and we considered only birds that colonise by chance events. We thus excluded marine and migratory species. We focused on songbird-like and pigeon-like birds, which constitute the majority of terrestrial (land-dwelling) birds on islands. We included only species from the same trophic level: we excluded aquatic birds, birds of prey, rails and nightjars. We also excluded introduced and vagrant species. Sex and age of the individuals is not relevant for the purposes of this study. The full lists of species and samples are given in Supplementary Data 1 and Supplementary Information Table 3.
Sampling strategy	The sample size is the number of island colonisation plus island speciation events (569). We sampled all taxa of our focal group on each of 41 archipelagos.
Data collection	We sampled DNA from birds for sequencing, and compiled published sequence and phylogenetic data available. Bird samples were collected in the field by M.M, B.H.W., S.M.C., J.C.I, C.T. and L.V. New sequences were produced by K.H. and J.C.I. GenBank data and published phylogenetic trees were compiled by L.V.. New phylogenetic trees were produced by L.V. . Data on island physical features were compiled from various published sources cited in Extended Data Table 1.
Timing and spatial scale	Field work was conducted between 1999 and 2017 on the island and continental regions specified in Supplementary Data Table 3. The spatial scale is global, as field locations were located in several continents and oceans.
Data exclusions	No data were excluded.
Reproducibility	The likelihood and simulation analyses conducted in this study can be reproduced using examples provided in the R package DAISIE. We provide examples of the code and the same data used for running these analyses (e.g. see examples at the end of DAISIE_sim_global and DAISIE_MW_ML functions in the DAISIE R package).
Randomization	This is not relevant as we did not conduct experiments.
Blinding	Blinding is not relevant, as we did not conduct experiments.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	We conducted fieldwork on several islands worldwide in order to collect DNA samples from birds. The field conditions varied, but field work was only conducted when it was not raining to avoid hurting birds. The exact field conditions are not relevant because they do not impact the results.
Location	The 41 archipelagos/islands sampled are: Aldabra Group; Ascension; Azores; Bermuda; Canary Islands; Cape Verde; Chagos; Chatham; Christmas Island; Cocos (Costa Rica); Cocos (Keeling); Comoros; Fernando de Noronha; Galápagos; Gough; Guadalupe; Hawaii; Juan Fernández; Lord Howe; Madeira; Marianas; Marquesas; Mauritius Isl.; New Caledonia; Niue; Norfolk; Ogasawara; Palau; Pitcairn; Rapa Nui; Reunion; Rodrigues; Saint Helena; Samoa; SãoTomé e Príncipe; Selvagens; Seychelles (Inner); Society; Socorro; Tonga; Tristan da Cunha. Mainland sample locations: Angola, Andalucia (Spain), Cameroon, Equatorial Guinea, Gabon, Madagascar, Morocco. All relevant parameters (area, isolation, latitude, longitude, elevation, age) are listed in Supplementary Data 2.
Access and import/export	Information on collecting and export permits: - Angola - Biodiversity Research Protocol ISCED-Huíla and the South African National Biodiversity Institute (SANBI) (M.M.). - Cameroon - Limbe Botanical and Zoological Garden, Ministry of Scientific Research, Ministry of Forestry and Wildlife (M.M.). - Cape Verde - Cape Verde Agriculture and Environment Ministry; Ref.: 10/10 and 18/2015 (J.C.I.). - Comoros - Centre National de Documentation et de Recherche Scientifique, 2000 (B.H.W.) - Equatorial Guinea - Universidad Nacional de Guinea Ecuatorial (M.M.).

- Gabon - Centre National de la Recherche Cientifique (CENAREST), Station de Recherche de l'IRET at Ipassa-Makokou, Parc de La Lekedi, CENAREST NºAR0053/12/MENESTFPRSCJS/CG/CST/CSAR 2012 (M.M.).
- Madagascar - Ministère des Eaux et Forets, 2002 (B.H.W.)
- Mauritius/Rodrigues - National Parks and Conservation Service (Republic of Mauritius), 1999 (B.H.W.)
- Mayotte - Direction de l'Agriculture et de la Foret, 2000 (B.H.W.)
- Morocco - Ref: 5061/08/HCEFLCD/DLCDPN/PRN/CFF // 14-2015 (J.C.I.)
- New Caledonia, Loyalty Islands - Direction du Développement Economique, 2 Dec 2011 6101-858/PR (L.V.); 31 Jan 2014 6101-43/PR (S.M.C.).
- New Caledonia, South Province - Direction de l'Environmement Province Sud, 21 Jan 2014 Province Sud 3177-2013/ARR/DENV (S.M.C.)
- Portugal - Regional governments of:
1) Azores: 12/2016/DRA (J.C.I.)
2) Madeira: 02/2016 FAU MAD (J.C.I.)
- São Tomé e Príncipe, Direcção Geral do Ambiente, Ministério das Obras Públicas, Infraestruturas, Recursos Naturais e Ambiente 1999-present (no number) (M.M.)
- Seychelles - Bureau of Standards and Ministry of Environment, Centre National de Documentation et de Recherche Scientifique, 2000 (B.H.W.)
- Spain - Regional governments of:
1) Andalucía: SGYB/AF/FJRH/RE-35-36/13 (J.C.I.)
2) Canary Islands: Ref.: 443/02-10-2012 // Ref.: 2016/811 (J.C.I.)
- Reunion – CRBPO (Centre de Recherches sur la Biologie des Populations d'Oiseaux, Muséum National d'Histoire Naturelle, Paris), #602, 2007 (C.T and B.H.W).
- Museum samples: Department of Ornithology and Mammalogy of the California Academy of Sciences (Laura Wilkinson & Maureen Flannery); Natural History Museum at Tring (Mark Adams); Stuttgart State Museum of Natural History.

Disturbance

Minimal disturbance to sites - we used mist-nets, which are placed temporarily and cause minimal impact.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Study did not involve laboratory animals.

Wild animals

Birds were caught in the field using mist-nets and immediately released in the same location after a blood sample was taken. No bird was injured, killed or kept captive.

The new samples collected for this study comprised 90 different species (252 individuals): Acrocephalus rodericanus; Agapornis pullarius; Alectroenas sganzini; Anabathmis hartlaubii; Anabathmis newtonii; Anabathmis reichenbachii; Chalcophaps indica; Chrysococcyx cupreus; Chrysococcyx lucidus; Coccyzus melacoryphus; Columba larvata; Columba malherbii; Columba thomensis; Coracopsis vasa; Corvus albus; Crithagra burtoni; Crithagra capistrata; Crithagra mozambica; Crithagra rufobrunnea; Crithagra sulphurata; Cyanolanius madagascarinus; Cyanomitra olivacea; Dicrurus ludwigii; Dreptes thomensis; Erythrura psittacea; Erythrura trichroa; Estrilda astrild; Estrilda melpoda; Estrilda astrild; Euplectes albonotatus; Euplectes aureus; Euplectes capensis; Euplectes hordeaceus; Euplectes orix; Euplectes albonotatus; Humblotia flavirostris; Lanius newtoni; Leptosomus discolor; Lonchura cucullata; Motacilla bocagii; Myiagra caledonica; Nesoenas picturata; Nigrita bicolor; Nigrita canicapilla; Ploceus cucullatus; Ploceus grandis; Ploceus insignis; Ploceus melanogaster; Ploceus nigerrimus; Ploceus princeps; Ploceus sanctithomae; Ploceus velatus; Ploceus xanthops; Prinia molleri; Prinia subflava; Progne modesta; Quelea erythrops; Quelea quelea; Saxicola torquata; Serinus albogularis; Serinus citrinelloides; Serinus citrinipectus; Serinus flaviventris; Serinus flavivertex; Serinus mozambicus; Serinus totta; Streptopelia senegalensis; Streptopelia decaocto; Sylvia atricapilla; Sylvia borin; Sylvia dohrni; Terpsiphone atrochalybea; Terpsiphone rufiventer; Terpsiphone rufocinerea; Terpsiphone smithii; Terpsiphone viridis; Treron calvus; Treron griveaudi; Treron sanctithomae; Turdus merula; Turdus olivaceofuscus; Turdus xanthorhynchus; Turtur afer; Turtur tympanistria; Uraeginthus angolensis; Vidua macroura; Zosterops feae; Zosterops griseovirescens; Zosterops leucophaeus; Zosterops lugubris.

Sex and age of the individuals is unknown (and not relevant for this study).

Field-collected samples	Blood samples collected in the field were stored in ethanol.
Ethics oversight	No ethical approval was required as no bird was killed, injured or kept captive and we used normal procedures for mist-netting.

Note that full information on the approval of the study protocol must also be provided in the manuscript.